(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 July 2003 (17.07.2003)

PCT

(10) International Publication Number
WO 03/058518 A2

(51) International Patent Classification[7]: G06F 17/60, A63F 13/12, H04L 29/06

(21) International Application Number: PCT/GB03/00031

(22) International Filing Date: 7 January 2003 (07.01.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0200255.8 7 January 2002 (07.01.2002) GB
0208146.1 9 April 2002 (09.04.2002) GB

(71) Applicant and
(72) Inventor: CRAMPTON, Stephen, James [GB/GB]; 9 Broadfields, Goffs Oak, Waltham Cross, Hertfordshire EN7 5JU (GB).

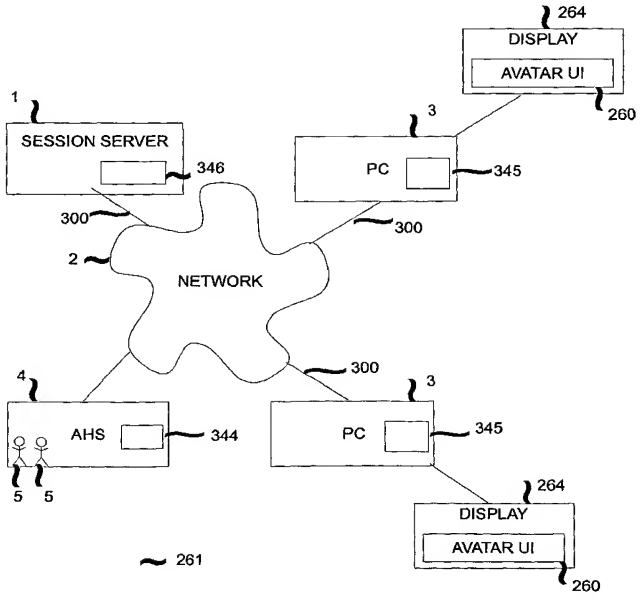
(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHOD AND APPARATUS FOR AN AVATAR USER INTERFACE SYSTEM



(57) Abstract: Apparatus for an Avatar User Interface System (261) is provided comprising Personal Computers (3), an Avatar Hosting Server (4) and a Session Server (1) all connected by a Network (2). Avatars (5) of individuals are retrieved from the Avatar Hosting Server (4) and displayed on a Display (264) in an Avatar User Interface Window (260). The Avatar User Interface System (261) is operable by individuals such that an individual using a Personal Computer (3) will see the Avatars (5) of other individuals on the Display (264) in an Avatar User Interface Window (260).

WO 03/058518 A2

METHOD AND APPARATUS FOR AN AVATAR USER INTERFACE SYSTEM

FIELD OF THE INVENTION

The present invention concerns methods and apparatus for an avatar user interface system to people, information, media and agents with photo-realistic avatars.

BACKGROUND TO THE INVENTION

It is well established in the marketplace that face to face communications have significant advantages over ways of communicating virtually such as video conference calls and audio conference calls. With the increasing globalisation of business and the shrinking timescales of new commercial initiatives, it is even more important to communicate well. But at the same time, the cost of travelling to face to face communication sessions is increasing.

An alternative method of communicating is in a virtual world. Several companies have provided 3D worlds with avatars including Blaxxun (Germany) with its consumer world Cybertown. In these worlds, the user navigates his avatar into proximity with one or more avatars and chat then commences involving the owners of the avatars. User-driven gestures are incorporated. The avatars used in these virtual worlds are not photo-realistic representations of the person they represent.

Photo-realistic avatars of people can be generated in Avatar Booths as disclosed in UK Patent GB 2336981. An ad hoc standards group called H-anim has drafted a version H-Anim 2001 for avatars that can be found on the world wide web at www.h-anim.org. These photo-realistic avatars are also becoming anima-realistic: they can be animated realistically. Harold Sun and Dimitri Metaxas published a solution to generating life-like walking animation for an avatar automatically following a path in the proceedings of SIGGRAPH 2001 p 261-269.

As well as moving anima-realistically, the avatars need to talk anima-realistically. Eric Cosatto and Hans Peter Graf in their paper 'Sample-based Synthesis of Photo-Realistic Talking Heads' given at the Computer Animation conference Jun 8-10 1998 in Philadelphia show a system with a talking head speaking from a synthesis of text. Their

paper explains the conventional approach of generating lip movements from phonemes and how co-articulation is handled.

The present invention aims to provide avatar user interface system means by which a user has a high sense of presence that overcomes some of the disadvantages of other communication methods. Embodiments of the present invention use photo-realistic avatars of the participants in the communication session to create a virtual communication room with high photo-realism and high anima-realism. Embodiments of the present invention provide an avatar user interface system in which a synchronous communication session can take place without the user needing to control the user interface manually and thus allowing the user to concentrate on communicating. Embodiments of the present invention provide an avatar user interface system in which multi-tasking can take place between multiple communication and information processing tasks. Embodiments of the present invention provide an avatar user interface system in which people and agents may communicate with each other.

There is a multitude of applications of an avatar user interface system. Some examples of significant commercial applications are given below.

- conferences
- meetings
- e-learning tutorials
- product presentations
- exhibitions
- call switchboard
- multi-tasking communication tool
- security
- interactive games
- collaborative work
- shared space virtual reality
- social exercise
- practicing

SUMMARY OF THE INVENTION

In accordance with one aspect of the present invention there is

provided an apparatus for an avatar user interface system comprising:

- server means for serving the communication session;
- one or more computing appliance means;
- network means for joining said server means and said computing appliance means;
- avatar means for representing each user visually; and
- avatar user interface application means resident on each computing appliance means;

operable by one or more users.

In accordance with this aspect of the present invention there is provided a method of communication between a plurality of users via an avatar user interface system comprising the steps of:

- joining a plurality of computing appliance means and a server means for serving the communications session to start a communication session by means of a network;
- viewing the avatars of the users involved in the communication session on the said plurality of computing appliance means;
- a user first communicating into a computing appliance;
- one or more users receiving the first communication on one or more other computing appliances;
- avatars enacting the first communication on said computing appliances;
- a user responding to the first communication in a second communication;
- one or more users receiving the second communication on one or more other computing appliances;
- avatars enacting the second communication on said computing appliances;
- continuing the exchange of communications until the session is finished; and
- terminating the joining of the computing appliance means and the server means for serving the communications session to terminate the communication session.

In accordance with a further aspect of the present invention there is provided a method of communicating between at least one user and at

least one avatar agent via an avatar user interface system comprising the steps of:

- joining one or more computing appliance means, an avatar agent hosting server means hosting one or more intelligent agent software units and a server means for serving the communications session to start a communication session by means of a network;
- viewing the avatars of the said avatar agents and said users involved in the communication session on the said computing appliance means;
- a user or an avatar agent first communicating;
- if there are one or more users who did not first communicate, then the one or more users who did not first communicate receive the first communication on one or more other computing appliances;
- avatars enacting the first communication on said computing appliances;
- if there are one or more avatar agents who did not first communicate, then the one or more avatar agents who did not first communicate receive the first communication;
- a user or an avatar agent responding to the first communication in a second communication;
- one or more users or one or more avatar agents receiving the second communication;
- if there are one or more avatars receiving the second communication, then avatars enact the second communication on said computing appliances;
- continuing the exchange of communications until the session is finished; and
- terminating the joining of the computing appliance means, the avatar agent hosting server means and the server means for serving the communications session to terminate the communication session.

In a further aspect, the present invention aims to provide an integrated multi-media communication system for use in a broad range of applications based around photo-realistic avatars for communication with people and intelligent agents in both synchronous and asynchronous ways that is supportive of multiple concurrent communication sessions and of switching between communication

sessions.

In a further aspect, the present invention aims to provide a user interface system in which avatar means may be photo-realistic avatar means or parameter avatar means or animatable image avatar means.

BRIEF DESCRIPTION OF THE DRAWINGS

Embodiments of the present invention will now be described, by way of example only, with reference to the accompanying drawings, in which:

Figure 1 is a block diagram of apparatus for an avatar user interface system in accordance with a first embodiment of the present invention;

Figure 2 is a schematic diagram of an avatar;

Figure 3 is a block diagram of avatar visual types;

Figure 4 is a block diagram for the reconstruction of a parameter avatar;

Figure 5 is an example table of avatar parameters;

Figure 6 is a block diagram of apparatus for generating and editing a parameter avatar;

Figure 7 is a list of action impersonation parameters stored in the memory of a personal computer;

Figure 7a is a flow diagram illustrating the process for defining action impersonation parameters and action impersonation rules for an activity;

Figure 8 is a block diagram of apparatus for generating and editing action impersonation parameters;

Figure 9 is a schematic diagram of an avatar hosting server system;

Figure 10 is a schematic diagram of an avatar number;

Figure 11 is a block diagram of a personal computer with an avatar user interface;

Figure 12 is a diagrammatic representation of avatar user interface functionality in an avatar conference application;

Figure 13 is a block diagram of a presentation media window;

Figure 14 is a block diagram of a whiteboard media window;

Figure 15 is a representation of an example of a meeting room media window;

Figures 16a, 16b, 16c and 16d are schematic diagrams to illustrate possible virtual camera positions in a virtual video conference;

Figures 17a, 17b and 17c are schematics of three possible layouts in the meeting room media window;

Figure 18 is a plan view of the virtual meeting room illustrating possible virtual camera positions;

Figure 19 is a set of four timelines of the camera shots during an avatar user interface session in four modes;

Figures 20 is a block diagram of a software director and avatar engine player;

Figure 21 is a block diagram of events on a personal computer and a session server;

Figures 22a, 22b, 22c, 22d and 22e are schematics of the five seating plans viewed by the five participants;

Figures 23 is a schematic of the audio mixer;

Figures 24 is a schematic of the audio mixer for multiple conversations;

Figure 25 is a block diagram of a lip sync generator;

Figure 26 is a timeline of a lip sync generator;

Figures 27a, 27b, 27c and 27d are diagrammatic representations of four lip sync animation types that can be used to animate a talking head;

Figure 28 is a flow diagram illustrating the steps involved in a lip sync generator;

Figure 29 is a flow diagram illustrating the steps in the passage of sound through an avatar user interface system;

Figure 30a is a spectrogram;

Figure 30b is a graphical diagram of a spectrum;

Figure 31 is a block diagram of the session server system;

Figure 32 is a block diagram of an apparatus for holding an avatar user interface session using voice and data networks in accordance with a second embodiment of the present invention;

Figure 33 is a schematic diagram of an animatable image in accordance with a third embodiment of the present invention;

Figure 34 is a schematic diagram of an animatable image avatar;

Figure 35 is a schematic diagram of a set of four state images for the jaw and mouth segment;

Figure 36 is a tree diagram of the hierarchy of animatable avatar image components;

Figure 37 is a schematic diagram of an animatable image generator;

Figure 38 is a schematic diagram of an apparatus for animatable image generation;

Figure 39 is a block diagram of an avatar user interface system with multiple formats of avatar;

Figure 40 is a schematic layout of an avatar user interface with attendee functionality;

Figure 41 is a schematic layout of an apparatus for a multi-party location in an avatar user interface system in accordance with a fourth embodiment of the present invention;

Figure 41a is a schematic of the 3D sound processing;

Figure 42 is a representation of an example of the displayed avatar user interface with switchboard functionality in accordance with a fifth embodiment of the present invention;

Figure 43 is a block diagram of a multi-session server system;

Figure 44 is a block diagram of a stand-alone avatar user interface system in accordance with a sixth embodiment of the present invention;

Figure 45 is a representation of an example of the avatar user interface system with extended exhibition functionality in accordance with a seventh embodiment of the present invention;

Figure 46 is a block diagram of an avatar agent hosting system and intelligent agent software in accordance with an eighth embodiment of the present invention;

Figure 47 is a block diagram of an apparatus for generating impersonation parameters;

Figure 48 is a block diagram of the avatar user interface system with extended security functionality in accordance with a ninth embodiment of the present invention;

Figure 49 is a block diagram of an avatar user interface system for interactive computer gaming in accordance with a tenth embodiment of the present invention;

Figure 50 is a schematic of an avatar user interface system for a six-sided cave in accordance with an eleventh embodiment of the present invention;

Figure 51 is a schematic of an avatar user interface system for two caves connected by a network;

Figure 52 is a schematic of an avatar user interface system comprising two exercise stations connected together by a network in accordance with a twelfth embodiment of the present invention;

Figure 53 is a schematic of the display of an avatar user interface system with an avatar virtual environment as the background in accordance with a fourteenth embodiment of the present invention;

Figure 54 is a schematic of a terminal of an avatar user interface system including motion-tracking cameras in accordance with a fifteenth embodiment of the present invention;

Figure 55 is a block diagram of apparatus for an avatar user interface system with multiple user devices;

Figure 56 is a schematic of a display device consisting of a display screen, an AVE projector and a Presentation projector in accordance with a sixteenth embodiment of the present invention;

Figure 57 is a schematic of a display device in which the AVE and Presentation projection means are combined into one physical unit;

Figure 58 is a schematic of a multi-density display device comprising an area of low density pixels and an embedded area of high density pixels;

Figure 59 is a schematic of an avatar user interface system with a mixed audience of avatars of virtual users at various locations and physical users;

Figure 60 is a block diagram of an apparatus for presentation preparation.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

FIRST EMBODIMENT

Figure 1 is a block diagram of an apparatus for an avatar user interface system 261 in accordance with a first embodiment of the present invention.

Avatar Conference application

The avatar user interface system 261 invention can be embodied in many applications. The avatar user interface system 261 is disclosed in this first embodiment embodied as an avatar conference application. An avatar conference is an example of a communication session on an avatar user interface system 261. Further embodiments disclose the avatar user interface system 261 invention embodied in different applications.

In this embodiment, the apparatus comprises two or more personal computers 3 with memory 345, display devices 264 and displayed avatar user interfaces 260 that are connected by a network 2 to a session server 1 with memory 346 using a standard avatar interface protocol 300 and an avatar hosting server 4 containing a plurality of avatars 5 and memory 344.

As will be described in detail below in accordance with the present invention avatars 5 representing the parties taking part in the avatar user interface session are stored on the avatar hosting server 4. The avatars 5 are transferred to the personal computers 3 across the network 2. The session server 1 mixes the voice streams from the personal computers 3 and returns them to the personal computers 3. The avatars 5 are displayed in the displayed avatar user interfaces 260 of the display devices 264 of the personal computers 3.

Avatars

Figure 2 is a schematic diagram of an avatar 5. The avatar 5 has an avatar identity 275 comprising an avatar number 8, a password 9 and a display permission flag 259. Associated with the avatar 5 are one or more types of data which may include: photo-realistic visual avatar data 340, animatable image avatar segment data 395, other visual image data 396, avatar parameters 230, impersonation parameters 325, biometric data 317, intelligent agent software unit 320, billing data 342 and personal data 341. The impersonation parameters 325 are of two types: voice impersonation parameters 331 and action impersonation parameters 332. Each set of data associated with the avatar 5 may be resident on different servers on the network 2 or servers on other networks that may be accessible via the network 2.

Figure 3 is a block diagram of avatar visual types. The visual component of an avatar 5 may be a 3D avatar 39 or an animatable avatar image 382 or another avatar type 239. There are two types of 3D avatar 39: a photo-realistic avatar 238 and a parameter avatar 232. An avatar 5 includes at least one of the photo-realistic visual avatar data 340 or the avatar parameters 230 or the animatable image avatar segment data 395 or the other visual image data 396 and any other or all of the other types of data.

An avatar 5 comprising at least photo-realistic visual avatar data 340 is referred to as a photo-realistic avatar 238. An avatar 5 comprising at least avatar parameters 230 is referred to as a parameter avatar 232. An avatar 5 comprising at least animatable image avatar segment data 395 is referred to as an animatable image avatar 382. An avatar 5 comprising at least either photo-realistic visual avatar data 340 or avatar parameters 230 is referred to as a 3D avatar 39. An avatar 5 comprising at least other visual image data 396 is referred to as another avatar type 239.

Photo-realistic Avatars

Photo-realistic visual avatar data 340 is a computer model that represents an individual taking part in the avatar conference. It is photo-realistic. When viewed by a person who knows the individual that it represents, that photo-realistic visual avatar data 340 will

be recognisable as a photo-realistic avatar 238 of the individual in the same way that a photograph of an individual is recognisable by a person who knows the individual as being a photograph of an individual.

In this embodiment, the photo-realistic visual avatar data 340 is a three dimensional (3D) computer model. The structure of the photo-realistic visual avatar data 340 is similar in terms of its components to the draft H-Anim 2001 standard. In this embodiment, the external shape of the photo-realistic visual avatar data 340 is represented by polygonal meshes totalling approximately 6,000 polygons. A generic avatar topology is used in which every photo-realistic visual avatar data 340 of every person has the same number of polygons, whether the person is tall or short, fat or thin, male or female. Texture mapping is used to position images of the avatar over the polygons so that the avatar can be rendered to appear like the individual it represents. The compressed size of the photo-realistic visual avatar data's computer model is typically between 200 and 900 Kbytes.

In this embodiment a subset of the full number of joints specified in h-anim is used; in particular, not all the joints in the back, the hands and the feet are modelled. If all the joints were used, there would be considerable extra computational cost for very little extra anima-realism of movement.

Parameter Avatars

Photo-realistic visual avatar data 340 can be quite large and, on lower bandwidth connections, it can take a long time to download. For the avatar conference to feel right to the user, a person's avatar should be seen when he is speaking, rather than just heard as a disembodied voice. Ideally, the avatar should appear in the avatar conference at the same time as a person joins the conference. If it is known who will be in the conference when the conference is organised, then photo-realistic visual avatar data 340 can be sent out in advance of the start of the conference. However, if someone joins the conference without any notice, then it is a purpose of this invention to use parameter avatars 232 that are very small and that will appear shortly after the person joins.

Figure 4 is a block diagram for the reconstruction of a parameter avatar 232. A set of avatar parameters 230 is sent to a personal computer 3 that enable a parameter avatar 232 to be constructed from a general database of avatar information 231. Avatar parameter 230 download assumes that there is a general database of avatar information 231 already downloaded at the personal computer 3 from which a parameter avatar 232 can be quickly generated from a small set of avatar parameters 230. The general database of avatar information 231 is downloaded the first time that an avatar conference is accessed on a personal computer 3 and remains for later avatar conferences unless it is deleted.

Figure 5 is an example table of avatar parameters 230 that can be used to define a parameter avatar 232 from a general database of avatar information 231. This set of avatar parameters 230 is typically in the range of 100 to 1,000 bytes in size but may be smaller than 100 bytes or larger than 1000 bytes and thereby be sent over the network 2 from the avatar hosting server 4 to the personal computer 3 very quickly. The parameter avatar 232 can also be assembled very quickly from the database 231 and the avatar parameters 230. In this way, an avatar of the new participant can be constructed quickly that would look like that person from a distance.

This parameter avatar 232 may be displayed until such time as the photo-realistic avatar 238 has been downloaded from the avatar hosting server 4 to the personal computer 3 at which point the parameter avatar 232 is automatically replaced with the photo-realistic avatar 238. The photo-realistic avatar 238 can be downloaded progressively, such that rather than a sudden change from a parameter avatar 232 to a photo-realistic avatar 238, the user sees a slow morphing from one to the other over a period of time. Progressive download can be implemented in many ways. One implementation might be to first download the geometry, then the joint positions, then the textures. A second implementation might download low-resolution textures followed by high-resolution textures.

It is possible to use a large set of avatar parameters 230 and the

power of each parameter is such that an extensive database 231 can be used to generate very life-like parameter avatars 232. The most distinctive part of a human is the face. Faces can be generated that are very close to the actual person's face from as little as 50 avatar parameters.

Avatar generation

Avatars and parameter avatars may be generated in several ways:

- a photo-realistic avatar 238 may be generated from photos of the user
- a parameter avatar 232 may be built up manually by the user without using photos of the user
- a parameter avatar 232 may be automatically generated from a photo-realistic avatar 238 of the user

Parameter avatar generated from photo-realistic avatar

Figure 6 is a block diagram of apparatus for generating and editing a parameter avatar 232. The parameter avatar 232 may be generated automatically or manually.

A set of avatar parameters 230 is automatically created from a photo-realistic avatar 238 of the person by a parameter avatar generator 233 with avatar editing software 234. There is enough information in a photo-realistic avatar 238 for the avatar generator 233 to be relatively simple to create for those skilled in the art. The parameter avatar generator 233 is shown resident on a personal computer 3 but may be resident on an avatar hosting server 4 or any other server or computer on the network 2.

Parameter avatar generated manually by user

If a user 17 has not yet had a photo-realistic avatar 238 made of himself, then he can quickly create a set of avatar parameters 230 for a parameter avatar 232 that is roughly similar to him by providing input into the parameter avatar generator 233. Parameter avatar creation in the parameter avatar generator 233 is by selection by the user 17 of a number of graphical alternatives such as hairstyles and by entry by the user 17 of data such as height.

In the situation where a new user without an avatar needs to join his first avatar conference as quickly as possible, it is imperative that it is possible to create a 'rough' parameter avatar as quickly as possible. In these situations, users are very impatient and the interaction in which the parameter avatar is created must be very efficient and fast. Typically, under time pressure, the user may be prepared to spend 30-60 seconds on this interaction. The interaction would normally be one of selection of options with a mouse click rather than typing in data. Later on, the user may go back and spend more time refining his parameter avatar. It is a purpose of this embodiment that there are two or more ways of generating a parameter avatar depending on the amount of time that the user has available.

Action impersonation parameters

It is a purpose of this avatar user interface system invention that action impersonation parameters may be used to characterise how a person moves. One of the objectives of a successful avatar user interface system invention is anima-realism. It is a first objective for an avatar to move anima-realistically such that a user who does not know the person whose avatar it is, thinks that the animation is realistic. It is a second objective for an avatar to move anima-realistically whilst impersonating the actions of the person whose avatar it is, such that a user who knows the person whose avatar it is, thinks that the animation is both realistic and typical of that person. Achievement of this second objective will eliminate any dissatisfaction by the user of seeing an avatar of someone he knows behaving uncharacteristically and enable a deeper sense of copresence from use of the avatar user interface system invention. This avatar user interface system invention achieves the second objective by using action impersonation parameters.

Figure 7 is a list of action impersonation parameters 332 stored in the memory 345 of a personal computer 3. Action impersonation parameters 332 include: walking 400, running 401, ambient motion whilst standing 402, ambient motion whilst sitting 403, gestures whilst talking 404, facial expressions whilst talking 405 and lip synchronisation whilst talking 406. In the example of the action impersonation parameter for gestures whilst talking 404, there are a

number of possible gestural animations (actions) that might be associated with this 'gestures whilst talking' action impersonation parameter 404. These could include: waving hands excitedly in a beat mode whilst talking and moving hands to time with the end of a sentence.

Action impersonation parameters 332 are not limited to the above characteristics, but may be extended to include any characteristics required in an application of this avatar user interface system invention. For the purposes of this disclosure, a reference to action impersonation parameters 332 will mean reference to either or both of: types of action impersonation parameter and action impersonation parameter values.

Values for action impersonation parameters 332 depend on the type of action and its definition. Values are set for action impersonation parameters 332 of a particular person in their avatar 5. Alternatively, values may be assigned as a set of action impersonation parameters 332 for a generic person in or with a context. Examples of sets of generic values might include:

- an Italian person
- a hyperactive person
- a person in a meeting
- a hyperactive Italian in a meeting

A context for generic impersonation parameters might be a communication context. Examples of communication contexts include: meetings, product presentations, virtual exhibitions, receptions, major conferences, security situations, interactive game playing, exercise and practicing.

Values may also be assigned for individual action impersonation parameters 332 that are characteristic of a style. An example is walking, where styles of walk can be defined such as a rolling gait, a mincing step etc.

It is a purpose of this embodiment to disclose a manual process for an appropriate activity 337 of (i) defining types of action impersonation

parameter 332 involving the analysis of video of one or more people undertaking the activity 337 and (ii) deriving action impersonation rules 333 as to the context and frequency of use of each type of action impersonation parameter 332 in the activity 337. In the disclosure of this manual process, the activity used by way of example is a meeting, but this embodiment is not limited to the activity of meetings and is applicable to most types of human activity.

Figure 7a is a flow diagram illustrating the process for defining action impersonation parameters 332 and action impersonation rules 333 for an activity 337. In the first step S1000, a significant corpus of videos 336 of meetings is recorded. Each meeting will typically require several video cameras 29 to synchronously record different participants at a sufficient resolution. Using a plurality of cameras 29 overcomes the problem of one camera not being able to image participants to a high enough resolution sitting all the way around a table. Meetings with different numbers of participants are recorded. Meetings with people from different cultures may be recorded. Meetings with people of different personalities may be recorded. A video corpus 336 of 20-50 hours is a typical size for an activity 337.

In the second step S1001, the corpus is processed by a trained person along a timeline. The actions of each participant may be related to a number of parameters such as status, activity type (speaking, listening, observing), speech content and emotion. The result is an annotated timeline 334 with actions of each participant related to the parameters.

In the third step S1002, the annotated timeline 334 is analysed to produce: (i) a type definition of each possible action impersonation parameter 332, (ii) a set of rules that can be incorporated in a finite state machine 333.

Figure 8 is a block diagram of apparatus for generating and editing action impersonation parameters for an avatar 5 of a particular person. Action impersonation parameters 332 may be set manually by providing input from the user 17 into the action impersonation generator/editor 335. The user 17 may be the particular person whose

avatar it is or someone else such as a friend, a family member or an expert providing a service. Action impersonation parameters 332 may be edited manually by providing input from the user 17 into the action impersonation generator/editor 335.

Individual action impersonation parameter setting in the action impersonation generator/editor 335 may be by manual selection by the user 17 of a number of high-level visual alternatives for each individual action impersonation parameter such as walking style and by entry by the user 17 of data such as whether a particular gesture is typically used.

In the situation where a new user without an avatar needs to join his first communication context such as an avatar conference as quickly as possible, it is imperative that it is possible to set up a 'rough' set of generic action impersonation parameters as quickly as possible. This can be achieved at the highest level by providing the user with a small number of pre-set generic action impersonation parameter sets to choose between. Examples include:

- passive
- active
- hyper-active

An alternative high-level way of setting action impersonation parameters quickly is to choose between pre-set action impersonation parameter sets according to culture. The user may choose between cultural characteristics such as:

- Anglo-Saxon
- Japanese
- Hispanic
- Italian

After personal action impersonation parameters 332 have been set in a high-level, generic way, they may be edited at a low-level where they can really be fine-tuned to the way a person moves. For instance a person may be hyper-active and use a characteristic gesture a lot but never use another gesture. By editing at a low-level, the action

impersonation parameters 332 may be refined such that a user who knows the person whose avatar it is, thinks that the animation is both realistic and typical of that person.

It is a purpose of this embodiment to disclose a manual process for defining a set of action impersonation parameters 332 for a particular person using an action impersonation generator/editor 335 involving manual input by a user 17. In the first step, the user 17 makes selections from a number of choices at a high level. In the second step, the user edits those selections at a lower level.

For automatic setting of action impersonation parameters, a video camera 29 may make video recordings 336 of a person carrying out a number of pre-defined actions. The action impersonation generator/editor 335 may automatically set the action impersonation parameters by automatic processing of the video recording. In this process, the emphasis is on replicating the particular person's style in actions that have different styles. The camera 29 may be mounted in a booth 18.

It is a purpose of this embodiment to disclose an automatic process for setting a set of action impersonation parameters 332 for a particular person using an action impersonation generator/editor 335. In the first step, video recordings 336 are made of a person carrying out a number of defined actions. In the second step, the action impersonation generator/editor 335 automatically analyses the video recordings 336 to generate a set of action impersonation parameters 332.

Action impersonation parameters may be set by a number of means in addition to those disclosed. For example, videos can be made of a person carrying out a number of tasks and an expert may study the video and set the action impersonation parameters.

The processes disclosed above for manually and automatically generating, setting and editing action impersonation parameters 332 define a number of methods by example. This aspect of the invention is not limited to the processes disclosed, but covers all processes

for manually and automatically generating, setting and editing action impersonation parameters 332.

Avatar numbering

Each avatar 5 has a unique avatar number 8. An avatar 5 may contain multiple visual avatar data including a photo-realistic avatar 238, a parameter avatar 232 and an animatable image avatar 382. When an avatar 5 is first created, it is allocated a unique avatar number 8. At any point thereafter, visual avatar data of different types may be added, deleted or edited.

Avatar access permission

The password 9 when used together with the avatar number 8 gives the user 17 access to change the avatar 5 including other types of data such as personal data 341. The display permission flag 259 if set by a user 17 with a password 9 and avatar number 8 gives permission to all other users 17 to use the avatar 5 for viewing purposes such as in a displayed avatar user interface 260 without need of the password 9. Access permissions are not limited in this invention to the password 9 and the display permission flag 259. A range of access permissions may be created for access to different types of data by different users.

Avatar Hosting Server

Figure 9 is a schematic diagram of an avatar hosting server system. The avatar hosting server 4 contains a database 6, avatar hosting management software 229, and avatars 5. In this embodiment, each avatar 5 has a unique avatar number 8 and a password 9. The avatar hosting server 4 may also contain one or both of billing software 237 and avatar generation software 222.

When the avatar hosting management software 229 on the avatar hosting server 4 receives a request 7 over the network 2 from a personal computer 3 for an avatar 5, then the avatar hosting management software 229 will check with the database 6 to see if the request is accompanied by a valid avatar number 8 and password 9. If the request 7 is valid, then the avatar hosting management software 229 will send the requisite avatar 5 to the personal computer 3 in such a form that

it can be changed. If the request 7 is not accompanied by a valid password 9, then the avatar hosting management software 229 will check to see if the display permission flag 259 is set for the avatar 5 with avatar number 8. If the display permission flag 259 has been set, then the avatar hosting management software 229 will send the requisite avatar 5 to the personal computer 3 in such a form that the avatar 5 can only be displayed and cannot be changed. If the request 7 is not accompanied by a valid password 9 and the display permission flag 259 is not set for the avatar 5 with avatar number 8, then the avatar hosting management software 229 will not send the requisite avatar 5.

Photo-realistic Avatar Generation

Photo-realistic avatars 238 of people are generated and edited from digital images 19 of people, usually taken from several sides of the person using a camera 221 using generation software 222 and avatar editing software 234 in an Avatar Generator Editor (AGE) 235.

The quickest and least technical means of generating these digital images 19 is by the person using a special avatar generation apparatus 18 such as an avatar booth run by generation management software 236. The generation management software 236 usually takes the images 19 of the person using a camera 221 and generates a photo-realistic avatar 238 using AGE software 235 on a personal computer 3. The special avatar generation apparatus 18 usually contains means for regulating the quality of the images 19 that reduces or eliminates the need for skilled processing of the images 19 before they enter the AGE software 235. Such regulation means usually include fixed camera settings, controlled lighting levels and a uniform colour and shape background and floor such as a chroma green sheet but neither have to include these regulation means or are limited by these regulation means.

Alternatively, any camera 221 can be used to take images 19 of the person in a largely unregulated way. These images can be transferred to a personal computer 3 on which AGE software 235 is resident. Alternatively, the images 19 can be sent over the network 2 to the avatar hosting server 4 on which there is also generation software 222 that automatically generates a photo-realistic avatar 238 without any

user intervention. Alternatively, the images 19 can be sent over the network 2 to an avatar generation service 223 that uses an AGE 235.

Avatar Generator Editor (AGE)

The automatic generation of an avatar or a parameter avatar generates an imprecise avatar. The avatar generated may not at first be pleasing to the user, in the same way that photographic images of a person are often not pleasing to the person. The user may think that the avatar does not represent himself or even his self-image.

In avatar and parameter avatar generation, an interactive editing process is possible to change the avatar. There are two main types of editing, both of which may be used:

- low-level: changing the avatar by touching up manually the 3D shape, textures, texture coordinates and joint positions
- high-level: changing the avatar by interactively adjusting avatar parameters from which the avatar is regenerated

It is a purpose of this embodiment to disclose an avatar generator editor (AGE) 235 containing a photo-realistic avatar generator 222 or a parameter avatar generator 233 and avatar editing software 234 in which editing can take place at low level or high level or both.

Peer to peer avatar serving

In an alternative to using an avatar hosting server 4, a peer to peer avatar serving system can be used. In a peer to peer avatar serving system, an avatar hosting server 4 is not required and the user's avatar 5 that is resident in local storage 274 on his personal computer 3 can be sent to all other participant's personal computers 3 directly over the network 2.

Avatar Hosting Services

Figure 10 is a schematic diagram of an avatar number 8. The avatar number 8 comprises two parts: an avatar hosting service identity number AHS-ID 224 and an avatar identity number A-ID 225. If there is multiple avatar hosting servers 4 on the network 2, then each avatar hosting server 4 has an avatar hosting service identity AHS-ID 224.

There is an avatar hosting registry server AHR 226 on the network 2 run by AHR management software 227 stored in memory 347. When a personal computer 3 needs an avatar 5 it takes the avatar hosting service identity AHS-ID 224 and sends it to the AHR management software 227 to request the location of the avatar hosting server 4 corresponding to the AHS-ID 224 on which the avatar 5 is stored.

Each avatar identity number 225 for a particular AHS-ID 224 is unique. The personal computer 3 contacts the avatar hosting management software 229 on the correct avatar hosting server 4 with the location provided by the AHR management software 227 and retrieves the avatar 5 using the AHS-ID 224.

It is a purpose of this first embodiment to disclose a process for retrieving an avatar comprising the following steps:

- user providing an avatar number and password;
- a computing appliance sends the avatar number and password to the network location of an avatar hosting service;
- avatar hosting server management software on the avatar hosting service checks a database to verify that the avatar number and password are valid;
- if the avatar number and password are valid, then avatar hosting server management software on the avatar hosting service sends the avatar to the computing appliance.

It is a further purpose of this first embodiment to disclose a process for retrieving an avatar using an avatar hosting registry server comprising the following steps:

- user providing an avatar number and password;
- a computing appliance sends an avatar hosting service identity number to an avatar hosting registry server;
- the avatar hosting registry server sends to the computing appliance the network location of the avatar hosting service corresponding to the avatar hosting service identity number;
- the computing appliance sends the avatar number and password to the network location of the avatar hosting service;

- avatar hosting server management software on the avatar hosting service checks a database to verify that the avatar number and password are valid;
- if the avatar number and password are valid, then avatar hosting server management software on the avatar hosting service sends the avatar to the computing appliance.

This invention is not limited to this one way of designing an avatar number 8 but includes all other ways of designing an avatar number 8 such that the avatar 5 with avatar number 8 may be located on one or more avatar servers.

Personal Computer

Figure 11 is a block diagram of a personal computer 3 with an avatar user interface 260 in an environmental location 273. The personal computer 3 includes a display device 264, a webcam 29, a headset 11 comprising microphone 12 and headphones 13, a keyboard 14 and a mouse 15 in a cabinet 16 running an operating system 20 which in this embodiment is the Microsoft Windows XP operating system, an avatar user interface software application 262 as a plug-in to the browser 263 in which the displayed avatar user interface 260 is seen by the user 17 in the browser window 21 on the desktop 423. The headset 11 is normally worn by the user 17 of the personal computer 3 during an avatar conference in such a way that the user 17 can hear through the headphones 13 and speak into the microphone 12. Each PC peripheral may be connected to the PC cabinet 16 by a wired or a wireless method; if it is a wireless method, the peripheral may contain a battery or be connected to a power source.

Information flowing

During an avatar user interface session (avatar conference call), those participating in the session will communicate via information flowing between the personal computers 3 and the session server 1. This information can be in different media formats including: voice, music, video, avatar animation, 3D models, presentation images, text, office application sharing, spreadsheets, word processor documents and whiteboard annotation.

Session server arrangement

The session server 1 may be resident on the network 2 in a server-client network design. Alternatively, the session server functionality may be resident on a personal computer 3 in a peer to peer network design. In this way, the personal computers 3 of the users 17 with session server functionality resident on at least one personal computer 3 is sufficient to use the avatar user interface system 261 over the network 2 without a separate session server 1.

Display arrangement

According to this embodiment, Figure 12 is a diagrammatic representation of avatar user interface functionality in a conference application. The personal computer 3 is running a personal computer operating system user interface 20 which is visible in the display device 264 as a desktop 423. The personal computer 3 is also running a network browser which is visible in the display device 264 as a browser window 21 and which in this embodiment is the Microsoft Internet Explorer browser Version 6. The personal computer 3 is connected over the network to the session server 1 via the browser window 21. The Uniform Resource Locator (URL) 22 active in the browser window 21 points to the session server 1. In the browser window 21 during a conference there is the avatar session user interface 10 comprising a large conference window 23, two smaller conference windows 24, 25 and one or more interaction windows 26. The large conference window 23 has control buttons 27; these buttons change depending on which media is being shown in the large conference window 23. An interaction window 26 has mode buttons 28.

The user interface may be 'always on' for the user 17 to speak. Alternatively, a button 272 is depressed by the user 17 when speaking and is acknowledged with the button 272 changing colour to show that the microphone is live. The button may also be activated by pushing a key on the keyboard 14.

The large conference window 23 is used to show whichever media is in use and requires the maximum resolution. The two smaller conference windows 24, 25 are for two other media formats.

The interaction windows 26 have several functions including: text chat, attendance list, address list, agenda and audio settings. The number of interaction windows 26 can be reduced by means of a window having several modes. In this embodiment there are two windows 26. The first window 26 is permanently dedicated to text chat. The second window 26 is controlled by mode buttons 28 for swapping between functions: attendance list, address list, agenda and audio settings.

The three conference windows 23, 24, 25 may have the same aspect ratio or may have different aspect ratios depending on the system design. The three conference windows 23, 24, 25 show the three avatar conference media windows: the presentation, the whiteboard and the meeting room. The user may select the media window in one of the two small conference windows 24, 25 to go into the large conference window 23 and the media window currently in the large window swaps back into the small window vacated by the selected media window.

Presentation media window

According to this embodiment, Figure 13 is a block diagram of a presentation media window 30 during an avatar user interface session. The presentation media window 30 can show images, slides, video clips and other visual media such as Flash from Macromedia Inc (USA) or applications such as computer games. The presentation media window 30 is controlled by the user using the control buttons 31 - 35, when it is in the large window, but cannot be operated when it is in a smaller window. There is a mode of use of the invention in which one party in the conference can make a presentation and the when the presenter changes a slide, the same slide will change in the presentation windows of all the parties.

Button 31 returns the presentation to the first slide. Button 32 moves back one slide. Button 33 moves forward one slide. Button 34 goes to the last slide in the presentation. Button 35 toggles between local control of the presentation and presenter control of the presentation.

Whiteboard media window

According to this embodiment, Figure 14 is a block diagram of a

whiteboard media window 40 during an avatar user interface session. The whiteboard 40 is controlled by sets of control buttons 41 - 43 when it is in the large window but cannot be operated when it is in a smaller window. The session server 1 maintains the whiteboard content as being identical on all client personal computers 3. The whiteboard consists of multiple pages on which content can be created or pasted. The analogy is that of a flip-chart which has multiple pages.

The set of control buttons 41 are similar in function to buttons 31 to 35 in the presentation window. They control which of the whiteboard pages is displayed. There can be local control of the whiteboard pages or control can be handed to the presenter by means of a mode toggle key.

The set of control buttons 42 presents a palette of colours for the person creating content to choose from. This is similar to the Microsoft Paint application.

The set of control buttons 43 presents a collection of tools for creating content. Examples include text mode, line drawing mode and rubout mode. These tools are similar to the Microsoft Paint application.

Meeting Room media window

According to this embodiment, Figure 15 is a representation of an example of a meeting room media window 50 during an avatar user interface session. There are 5 participants on the session. Each participant in the avatar user interface session is represented by their avatar 5 sitting around a meeting table 51. In the background is a screen 53 on which presentation slides are displayed, a whiteboard 54 which can be written on by the participants and the room comprising walls 55, ceiling 56, floor 57, door 58 with a door handle 59 and a windowpane 60. The avatars 5 shown in the meeting room media window 50 are labelled Ted, Jill, Andy and Pam. The avatar 5 labelled Pam is using a mobile phone 79. The avatar of Bert is not shown in Figure 15. Bert is viewing the meeting room media window 50 and is the fifth participant on the session. Bert does not see an avatar of himself. There may be other items in the room such as plants, sky

visible through the windowpane 60, birds flying in the sky and trees visible through the windowpane 60.

The user 17 arranging the conference may select from several designs of meeting room 50 offered by an avatar conference service provider. A selected meeting room 50 may be informal or formal. It may be large or small. It may be designed to suit a particular culture eg Japanese.

The buttons 45-48 control the mode 84 in which the meeting room media window operates. Button 45 selects mode M1. Button 46 selects mode M2. Button 47 selects mode M3. Button 48 selects mode M4. The layout button 85 controls the layout for modes in which the layout is an option.

The meeting room media window in an avatar user interface session is useful to different people at different times:

- if you have never physically met a person who is on the session, it is usually interesting to see their avatar to see what they look like
- when you come into the session, it is useful to visualise who is already there by seeing their avatars
- when someone arrives at or leaves the session, you can see who it is without the session being interrupted
- if you do not recognise the voice of the person speaking, you can see their avatar and their name label in the window

Video Conference Metaphor

In a video conference from multiple locations, there is usually either a split screen with a separate section in the monitor for each location or a separate monitor for each location. Many people have taken part in video conferences and are used to the Video Conference metaphor in which each location and often each participant are seen in a separate display section.

The main visual drawbacks of video conferencing are:

(a) that there is not a cohesive space for the meeting - each window is unrelated to the others

(b) that the patterns of gaze of the participants as seen in the monitors are disjointed; each participant tends to look in a different direction; this is at its worst in desktop video conferencing when webcams situated on top of personal computer monitors are used and the participant looks at the monitor and not at the webcam; this is unlike a real meeting in which the gaze of each participant has a function and there is a cohesive whole.

These two drawbacks significantly reduce the sense of copresence that video conferencing might offer and make the experience of participating in a video conference unsatisfactory. The concept of copresence has arisen comparatively recently and as yet there seems to be no commonly accepted definition of it. However, there is general agreement that where a high sense of copresence is experienced by users of the virtual environment, there are benefits varying from greater task efficiency to less distraction.

According to this embodiment, Figures 16a, 16b, 16c and 16d are schematic diagrams to illustrate the virtual camera positions in the virtual video conference. It is a plan view. Cameras 61, 62, 63 and 64 view avatars 5 labelled Ted, Jill, Andy and Pam respectively. Behind avatars 5 are four backgrounds 65, 66, 67 and 68.

According to this embodiment, Figures 17a, 17b and 17c are schematics of three possible layouts in the meeting room media window 50. Layout 1 shows the avatars 5 in a virtual room 69 sitting around a virtual table 51. Layout 2 shows the avatars 5 in a straight line arrangement. Layout 3 shows the avatars 5 in a split screen arrangement. The backgrounds 65, 66, 67 and 68 may be identical, similar or completely different depending on what works best for the selected Layout 1, 2 or 3. The layout is selected using layout button 85.

Meeting Room Metaphor

The Meeting Room media window 50 of the avatar conference is a metaphor for an actual meeting that is being video-cast live. An example might be a group discussion broadcast from a television

studio. By using photo-realistic 3D avatars, a photo-realistic 3D meeting room, anima-realistic animations of the avatars and good camera direction, it is possible to suspend the disbelief of the viewer on the session such that he thinks it is an actual meeting where he is the only person who is not in the room. This gives the viewer a higher sense of copresence in the avatar user interface session than is obtainable in a telephone conference call. The objective is for the enactment to be so realistic that the viewer finds it hard to tell the difference between the avatar conference and a live video of the actual meeting room.

According to this embodiment, Figure 18 is a plan view of the virtual meeting room illustrating possible virtual camera positions. Camera 71 is the overview camera and will show the view illustrated in Figure 15. Camera 71 is positioned at the eye position of the Avatar called Bert who is seeing the Meeting room media window 50 in Figure 15 on his personal computer 3. Cameras 72, 73, 74 and 75 view avatars 5 labelled Ted, Jill, Andy and Pam respectively. Camera 76 shows the presentation screen 53. Camera 77 shows the whiteboard 54. Other cameras may be positioned at any location and oriented at any orientation.

Meeting Room media window modes

There are four modes M1, M2, M3, M4 for the Meeting Room media window. The user is free to select a preferred mode using the buttons 45-48.

In each mode, the view presented is from a virtual camera position. In each mode there are one or more virtual cameras. A virtual camera can have camera controls such as zoom and pan in addition to spatial movement.

According to this embodiment, Figure 19 is a set of four timelines of the camera shots during the avatar conference for each Mode. In Mode M1, by way of example, there is only one shot S1 which lasts for the duration of the avatar conference and is shot from Camera 71. In Mode M2, by way of example, the first shot S10 is from Camera 71 and is an overview view similar to that in Figure 15. This is followed by shot S11 from Camera 72 which shows Ted. This is followed by shot S12 from

Camera 76 which shows the presentation screen. The avatar conference timeline continues until the last shot S17 from Camera 71. In Mode M3, by way of example, there is only one shot S20 which lasts for the duration of the avatar conference and is in Layout 1 using Cameras 61, 62, 63 and 64. In Mode M4, by way of example, the first shot S30 is in Layout 1. This is followed by shot S31 from Camera 61 which shows Ted against background 65. This is followed by shot S32 from Camera 76 which shows the presentation screen. The avatar conference timeline continues until the last shot S37 in Layout 1.

M1 Meeting room: Overview

This mode M1 uses the Meeting Room metaphor. An overview from a single virtual camera of: the table 51, all the avatars around it 5, the whiteboard 54 and the presentation screen 53. There are no other cameras.

The viewer's avatar is not present. If the viewer's avatar were present, then the viewer sees his own avatar animating and in particular lip syncing whilst he talks and the effect would be like a mirror that reflects actions you do not make. Seeing your own avatar breaks the metaphor and reduces the copresence felt by the viewer. The camera viewpoint can be from where the viewer could be sitting at the table or any other viewpoint that 'misses' the viewer's avatar.

M2 Meeting room: Chat show

This mode M2 uses the Meeting Room metaphor. Multiple cameras are used but there is only one window. The result is like a televised chat show with cuts from one camera to another as the chat develops.

M3 Video conference: Overview

This mode M3 uses the Video Conference metaphor but improves on it to partially overcome the drawbacks of a real video conference. The Meeting Room media window 50 is laid out in sections and shows one participant's avatar in each section of the window. Referring again to Figure 15, one of the three layouts in Figures 17a, 17b and 17c can be chosen by the user by toggling button 85. In all of the three layouts, the gaze direction of the avatars can be controlled to

overcome the drawback of a video conference or desktop conference in which the gaze direction of the avatars is disconcerting to the user.

Layout 1 helps to give a sense of cohesive space for the video conference in that the layout is enhanced with a virtual room 69 which can include all items shown in Figure 15 including a virtual table 51.

Layout 2 goes half way to providing a sense of cohesive 3D space by putting the avatars in a line but does not include a virtual room and virtual table.

Layout 3 is a split screen layout that maximises the display resolution per participant and is useful where there are a large number of participants in the avatar conference.

M4 Video conference: Chat show

This Mode M4 uses the Video Conference metaphor. Multiple cameras are used but there is only one window. The result is like a televised multi-location show with each participant in a different location with cuts from one camera to another as the chat develops.

Activities during an Avatar Conference

The avatar user interface system is used in a variety of ways. The following is a list of collaborative meeting activities and the percentages are an indication of the % of meeting time devoted to each activity type when averaged over a wide variety of meeting types.

- Discussion (no media) 59%
- Presentation (slides, images ...) 27%
- Discussion (white board) 5%
- Shared application (eg Word, Excel) 4%
- Watch video clip 3%
- Listen to audio clip 1%
- View 3D virtual object 1%

In addition to the collaborative meeting activities, individuals or small groups can perform other activities. These include:

- Text chat
- Whispering (private voice connection)
- Break-out meeting
- Preparing a whiteboard sheet
- Using an on-line translation service
- Multi-tasking with non-meeting activity eg reading, doing e-mail

Conference types

There are a variety of different conference types, agendas and objectives. Designers may discuss a 3D object. Advertising people may listen to radio adverts or view prototype packaging images, video clips of TV adverts. Businessmen may view a slide presentation. Salesmen may present new products. Students may take part in an e-learning course led by a tutor or they may work collaboratively together.

Different user interface displays

It is a purpose of this embodiment that the graphical display of the avatar user interface varies according to the computing appliance capabilities, the type of conference being held and user preference. Figure 12 is just one example of an avatar conference display. This invention is not limited to the one example shown in Figure 12.

Events during an Avatar Conference

The avatar conference is a series of events. The events are largely un-scripted, although there is often an agenda and a Chairman whose objective is to ensure that the meeting follows the agenda. The following events are listed by way of example only and do not form a comprehensive list of all events that can take place in an avatar conference:

1. Person joins the conference
2. Person leaves the conference
3. A person stops speaking
4. A person starts speaking
5. Two or more people speak simultaneously
6. A presentation slide is projected

7. The presentation projector is turned off
8. A video is shown
9. A new whiteboard sheet is drawn on
10. A previous whiteboard sheet is turned to
11. A camera shot times out and a new camera shot begins
12. Move onto a new agenda item
13. Write a minute of a point just discussed

Most of these events are generated as a result of the action or inaction of a participant in the conference as detected by input mechanisms such as keyboard, mouse and microphone into the avatar conference system.

Software director

According to this embodiment, Figure 20 is a block diagram of a software director 80 and an avatar engine player 210. The flow of events 81 into a software director finite state machine 80 is shown with the resulting flow of camera shots 82, light settings 214 and actions 83 such as avatar animations into an avatar player engine 210. The avatar player engine 210 also uses at least one avatar 5, the scene 211, props 215 and the lighting model 212 to combine with the shots 82, light settings 214 and actions 83 to generate and display the avatar conference on the avatar session user interface 10. A 3D graphics processor chip 213 is often used in the personal computer 3.

Since no physical meeting room exists, the avatar conference can be enacted with each event being acted out by an avatar. The enactment of the avatar conference can be shown from multiple camera viewpoints and camera movements such as translation, zoom and pan.

It is a purpose of this embodiment of the invention that a software director 80, which is a finite state machine, directs the enactment and visualisation of the meeting in the avatar conference media window by reacting to the events 81 as they occur during the meeting.

The software director 80 takes into account the mode 84 and layout 85. A library of actions 87 is available. An action generator 88 is available. These actions are animations for avatars. Action

impersonation parameters 332 from at least one avatar 5 are available. In addition, timers 86 are started after some actions and new events are triggered by timers 86 expiring.

The software director finite state machine 80 is effectively a software agent that initiates actions triggered by events according to rules. In a constrained activity such as an avatar conference, it is quite feasible to completely define all the events, all the actions and the set of rules for actions being generated by events.

In addition to fixed rules, some actions are generated randomly. The generation of random actions such as camera cuts and avatar gestures can make the avatar conference more realistic and less predictable to the viewer.

In generating actions for an avatar 5, the software director 80, takes into account the action impersonation parameters 332 of that avatar 5. In this way, the actions 83 generated for that avatar 5 can be more characteristic of the user 17 that the avatar 5 represents.

For example, if the action impersonation parameter 332 for gestures whilst talking 404 indicates a lot of gestures, then the software director 80 will generates actions 83 involving a lot of arm movement. In a similar way, if the action impersonation parameter 332 for lip synchronisation whilst talking 406 indicates very little lip movement whilst talking, then the software director 80 will generates actions 83 for lip synchronisation involving very little talking. Rules for the five other disclosed action impersonation parameters 332 [400, 401, 402, 403 and 405] may be drawn up in a similar way and for any other action impersonation parameters 332 that are defined and used.

It is also possible to use other software agent approaches to make the conference realistic; one example is fuzzy logic.

Animation player engine

The scene 211 is typically that of a room as illustrated in Figure 15. Each item in the scene is modelled in 3D. To achieve a close to video experience that encourages a sense of presence, each item is made of

photo-realistic textures as well as a 3D topology. Props 215 are 3D items in the scene that can be moved by the avatars or under self-power. Props 215 are modelled in a similar way to the scene.

A lighting model 212 is used. The light levels 214 of the lights in the lighting model 212 can be changed by the software director 80 in reaction to events during the avatar conference.

The visual aspect of the avatar conference is a collection of 3D content including multiple avatars, a scene, props and a lighting model. If rendering effects such as shadows are required, the complexity increases. This can provide a large load on the personal computer 3. More and more often, a powerful 3D graphics processing chip 213 is built into the personal computer 3. In this way, it is possible for the avatar conference to achieve an acceptable frame rate such as 15-25 frames per second.

Event accumulator

According to this embodiment, Figure 21 is a block diagram of events on a personal computer 3 and a session server 1. It illustrates the event accumulator 89 on the session server 1 that gathers events 81 and sends the accumulated events 81 to the software director 80 on each personal computer 3 via a network 2. A software director 80 can also generate events 81 and send them to the event accumulator 89.

The event accumulator 89 on the session server 1 receives events 81 from a variety of sources including:

- Software director 80 on personal computers 3
- Text chat software
- Agenda manipulation software
- Login software
- Slide presentation software
- Whiteboard software
- Lip sync generation

The session management software 228 manages one or more user interface sessions on the session server 1.

Session and Hosting Payment

Referring again to Figure 9, billing software 237 on the avatar hosting server 4 monitors aspects of the use of avatars such as the number of avatars hosted for a customer and arranges billing according to the revenue model agreed with the customer. As is appreciated by those skilled in billing, the billing software 237 is not limited to the functionality described above. For instance, the billing software 237 could monitor other aspects of the sessions, it could apply different revenue models to different customers, it could use micro-payments for immediate debiting during use, it could combine billing for sessions, billing for avatar hosting, billing for other services and it could be resident on any computer or server.

Plurality of meeting room arrangements and enactments

According to this embodiment, Figures 22a, 22b, 22c, 22d and 22e are schematics of the five seating plans viewed by the five participants in the avatar conference in their five meeting room media windows 50. The table 51 and the presentation screen 53 are the same in each view. Each participant's avatar is abbreviated to the first letter of its name: B for Bert, T for Ted, J for Jill, A for Andy and P for Pam. In each view one avatar is not shown: the avatar of the viewer. In effect the seating plan is rotated with reference to the presentation screen 53 for each of the five views. Each meeting room arrangement is therefore different. Other seating arrangement rules can be drawn up.

Since the arrangements are different for each viewer, the enactment's will also be different. If Ted's avatar enters the virtual room through the door 58, then in Figure 22a, Bert's view, he will have to walk to the far chair and sit down, whilst in Figure 22d, Andy's view, Ted will sit down at the chair nearest him.

Each viewer only sees his representation of the virtual meeting room and does not see the representations of the virtual meeting room on other participant's personal computers 3. The meeting participants will not be aware of this unless they hold discussions along the lines of "Jill, who is sitting on your left." Therefore there should not be

any confusion stemming from different meeting room arrangements and enactment's.

It is a further purpose of this embodiment that each meeting room window 50 displayed on each Personal Computer 3 can show a different representation of the virtual meeting room and a different enactment of the meeting.

Audio mixer

According to this embodiment, Figure 23 is a schematic of the audio mixer 90. It illustrates the audio mixer 90 that is part of the session server 1 and includes a balance system 204 and a filter system 205. N audio input streams 91 arrive at the session server 1 from the personal computers 3 over the network 2. One audio input stream 91 arrives from each personal computer 3. In addition, one or more audio input streams 92 might be available; audio input streams 92 can be generated from playing a media object during the avatar conference such as an audio or video clip or as streaming media channels coming in over the network 2; an audio input stream 92 might be voice, music, radio, TV or any other audio stream. N audio output streams 93 are generated by the audio mixer 90 and sent to the N personal computers 3 over the network.

The audio mixer 90 is a finite state machine that follows one main rule in the case of a conference where there is a single conversation common to all participants: the audio output stream 93 going to a personal computer 3 is a mix of one media object audio stream 92 and all the input audio streams 91 from the other personal computers except for the one coming from that personal computer. Audio mixing in the audio mixer 90 is digital and as will be clear to those skilled in the art is carried out by combining synchronised time segments such that the real time of each input segment from each participant is the same.

Amplitude balancing

The audio mixer is also able to carry out an amplitude balancing function using the balance system 204 by balancing the amplitudes of the input audio streams 91 by reducing the amplitude of loud audio

streams and increasing the amplitude of quiet audio streams before mixing. In this way participants do not need to concentrate hard to hear quieter participants and do not get shocked by louder participants.

Audio filtering

The audio mixer is also able to carry out a filtering function using the filter system 205 filtering the input audio streams 91 to reduce annoying sound artefacts generated by the mixing process or by lags in the network 2. In this way participants enjoy a cleaner and higher quality audio experience during the conference.

Whispering, groups and other sessions

It is often the case in a conference that the meeting splits into smaller groups, each of which hold a separate conversation.

It is also a desirable feature that two or more people can whisper together whilst the main conference conversation proceeds without distracting the other participants. This is a case where the functionality of an avatar conference can be superior to that of a physical conference in which people whispering together is often a distraction to the other participants. In a physical conference, participants who are whispering can hear both the main conference conversation and their whispered conversation.

Visual feedback in the meeting room media window 50 can be provided to participants showing who is whispering and who has split into a smaller group. A simple way is for the avatars 5 of those whispering to automatically get up and move to the back of the room where they can be seen chatting together by others (but not heard). The same approach of forming a standing group can be used for small groups. For a formal break-out group, another meeting room can be used. So as not to lose visual continuity, the additional meeting room can be situated behind the wall 55 which can be made of glass like a large window 60 and the avatars in the additional meeting room can be visible through the glass wall 55.

For the case where a user 17 is involved in a session completely separate from the conference, this can be represented by his avatar 5 using a mobile phone 79. This conveys to the other participants that a user 17 whose avatar 5 is holding a mobile phone 79 does not have his full attention on the conference.

Multiple conversations

According to this embodiment, Figure 24 is a schematic of the audio mixer 90 for multiple conversations. It illustrates the audio mixer 90 that is part of the session server 1 when more than one conversation is taking place simultaneously during the conference. There are 3 conversations taking place: Conversation1 201, Conversation2 202 and Conversation3 203. Conversation1 201 in the CONV1 mixer 94 uses the input and output streams 1, 2 and 3. Conversation2 202 in the CONV2 mixer 95 uses the input and output streams 4 and 5. Conversation3 203 in the CONV3 mixer 96 uses the input and output streams 6, 7 and 8. The mixed output 97 of mixer CONV1 94 is also fed into the CONV2 mixer 95. The CONV2 mixer 95 is set up to combine conversation1 and conversation2 such that the output streams 4 and 5 include both conversation1 201 and conversation2 202 but the output streams 1, 2 and 3 do not include any element of conversation2 202.

It is a further purpose of this embodiment that the audio mixer 90 can be configured to support two or more conversations simultaneously. In addition, it is possible to combine the main conference conversation with whispering such that two conversations can be heard simultaneously.

It is also possible to combine a conversation with a digital audio stream 92 playing for example music so that both the music and the conversation can be heard simultaneously.

Lip Synchronisation Generator

According to this embodiment, Figure 25 is a block diagram of a Lip Sync Generator (LSG) 100 in which the microphone 12 receives voice 270 from a user 17 and background noise 271 from an environmental location 273. The resulting analogue audio stream 103 generated by the

microphone 12 is processed by a standard sound card 102 such as a Sound Blaster from Creative Technologies Inc (USA) that is in the personal computer 3. The digital output from the sound card 104 is input into the LSG 100 which first reduces the background noise 271 with a filter 205 and then outputs a stream of geometric positions 101. In addition, a digital audio transform stream 105 is output from the LSG 100. The digital audio transform stream 105 can also be the same as the input audio stream 91 to the audio mixer 90. A stream of events 81 is also output by the LSG 100 which travels over a network 2 to the event accumulator 89 on the session server 1.

According to this embodiment, Figure 26 is a timeline of a lip sync generator. It illustrates that the processing in the LSG takes time and the output 101 lags the input 104 by time T milliseconds.

Lip sync animation types

According to this embodiment, Figures 27a, 27b, 27c and 27d are diagrammatic representations of four geometric values for four lip sync animation types that can be used to animate a talking head 111 with a mouth 112. In Figure 27a, the jaw rotation angle B is the angle between the jaw 107 and the upper teeth 106 first geometric value that can be output from the LSG. In Figure 27b, the mouth length L is the distance between the two corners of the mouth 109. In Figure 27c, the lip rotation angle A is the angle between the angle of the teeth 106 and the angle of the lip 108. In Figure 27d, the tongue protrusion length P is the length of protrusion of the tongue 110 from its rearmost position.

Voice processing

In this embodiment, the microphone records sound from the person 17 speaking in the conference. Human voice is typically audible in the range 20 Hz to 20 kHz. The analogue signal 103 from the microphone 12 is processed to produce a digital audio stream 104 sampled at 16 kHz and 16 bits resolution by the sound card 102 in the personal computer. Sampling at 16 kHz and 8 bits was tried but the data was too sparse to allow the LSG to perform well in this particular avatar conference configuration. The output 101 from the LSG 100 is four real numbers,

one for each geometric value of a lip sync animation type at a sample rate of 30 per second.

According to this embodiment, Figure 28 is a flow diagram of the process followed by the LSG 100. The digital audio stream data 104 flows into a buffer 120. At regular intervals, a discrete Fourier transform 121 is performed on the audio data accumulated in the buffer 120 and a spectrum 146 is output. The spectrum 146 comprises a finite number of bins representing frequency ranges with the degree to which each bin is filled defining the amplitude of that frequency range. A jaw rotation analyser 123 outputs a value representing the jaw angle 124. A mouth length analyser 125 outputs a value representing the mouth length 126. A lip rotation analyser 127 outputs a value representing the lip angle 128. A tongue protrusion analyser 129 outputs a value representing the tongue protrusion 130. One or more emotion analysers 135 output strengths of emotion 136. The stream of spectrums 146 generated is the audio transform stream 105. The combination of real numbers 124, 126, 128 and 130 in a stream is the geometric position stream 101 in which one or more strengths of emotion 136 are included. The audio transform stream is compressed 131 to produce a compressed audio stream 132 and this is combined 133 with the geometric positions 101 to form a stream of packets 134 for transfer over the network 2 to the session server 1.

In order that lip sync animation can take place on any head, the geometric values 124, 126, 128 and 130 for the four lip sync animation types are each normalised and output by the respective analysers 123, 125, 127 and 129 in the range 0 to 1.0.

The LSG 100 operates at 62.5 Hz in that a discrete Fourier transform is performed on the digital audio stream data 104 accumulated during the previous 0.016 sec. The frequency spectrum is divided into 128 bins representing frequency ranges. The packets sent over the network are sent at a frequency of 30 Hz. Operation at rates in excess of 100 Hz was tried, but the LSG quality deteriorated due to a reduction in signal. These values are settings at which the LSG 100 works, but this invention is not limited to these precise settings and includes all settings that work for this process.

Audio compression

The audio compression 131 in which the stream of spectrums 105 is further compressed can be carried out by any of the compression-decompression routines known to experts in the field.

Many methods of compression of audio streams include carrying out discreet fourier transforms as one step in the compression. It is an advantage of this invention that a single discreet fourier transform is used for two purposes: lip synchronisation generation and audio compression. This invention requires less of the personal computer processing power than methods in which lip sync generation and audio compression are performed in separate processes.

Audio lag

According to this embodiment, Figure 29 is a flow diagram illustrating the steps involved in the passage of a sound from the microphone 12 on one personal computer 3, through the sound card 102, processed by the LSG 100, sent to the session server 1 over the network 2, buffered and mixed in the audio mixer 90, resent over the network 2 to another personal computer 3, buffered and decompressed 140 and played on the headphones 13 via the sound card 102.

The geometric and audio information in the packets is for the same period of time; in other words there is no lag within the packet between the geometric and audio information. This has the advantage of perfect timing on the lip synchronisation on replay. There is also the advantage of simplicity in which the two data types are combined in the same packet.

But there is a lag for the whole system in that, as shown in Figure 29, the sound passes though many stages from when it is spoken into the microphone 12 to when it is heard in headphones 13. The largest element of lag may be a different element in each system design. If the network 2 is the internet then the lag caused by the internet could be in excess of 1 second. The greatest acceptable lag in tele-conversations is around 200-300 milliseconds although longer lags of 500 milliseconds or more are considered acceptable by users on mobile

phone networks. The lag in the LSG between the digital audio input 104 and the outputs 101 and 105 is typically in the range T = 0.05 to 0.1 sec.

Geometry from spectrum

According to this embodiment, Figure 30a is a spectrogram 145 and Figure 30b is a graphical diagram of a spectrum 146 from time t on the spectrogram 145. The spectrum 146 comprises a finite number of bins representing frequency ranges with the degree to which each bin is filled defining the amplitude of that frequency range. For clarity in disclosing this embodiment, the spectrum 146 is segmented into just 7 bins corresponding to rows f1 to f7. As already disclosed, the number of bins in a spectrum 146 is likely to be much higher. The row f1 corresponds to the lowest frequencies collected and the row f7 corresponds to the highest frequencies collected, with f2-f6 covering frequency ranges in between. For clarity in disclosing this embodiment, the amplitude of each bin is split into ranges a1 to a6. The range a6 is the largest range of maximum amplitudes. When displayed on a colour screen, the spectrogram 145 can be depth encoded in discrete colours such that the square in row f1 is coloured with a colour signifying amplitude a1, the square in row f2 is coloured with a colour signifying amplitude a4 and so on for rows f3 to f7 of the spectrum 146. In practice, the amplitude is likely to be stored as a floating point number and only split into amplitude ranges for the purposes of visualisation on the colour spectrogram 145.

It was appreciated that the generation of geometric values for facial animation during speech is not a perfect science. Each person's voice pattern is unique and so is their facial animation during speech. For a real-time LSG to be useful, it must generate facial animation that is acceptable to the user.

It was also appreciated that restricting the facial animation to four geometric parameters was a simplification reducing the representation of something as complicated as movements of a human face during speech to four values. It was foreseen that if the system worked well for four geometric parameters, then it might be improved by adding further

geometric parameters and that each parameter may be defined differently from the definition disclosed here.

The approach taken to generating acceptable facial geometry for four geometric parameters from a single spectrum was an experimental and analytical approach. Software visualisation tools were developed for showing the colour spectrograms of voice as it was recorded through a standard microphone supplied with a low-cost headset. In the end a range of 64 amplitudes with 64 colours were chosen for visualising the shape of utterances on the spectrogram and used in the rules for determining geometry from the spectrum. Internally, the amplitude is a floating point real number.

Most work analysing voice has been by researchers coming from the voice recognition or voice synthesis communities. Their approaches have been strongly linked to concepts such as phonemes, visemes, diphones and co-articulation. As seen in the disclosure of this patent, the LSG 100 has a direct route between voice spectrum and the geometry output without attempting to go through intermediate concepts such as phonemes, visemes, diphones and co-articulation.

Other LSG attributes

One requirement for the LSG is for it to scale to non-speech utterances such as singing and laughing. The need is for the avatar to visually represent those utterances in an acceptable manner. Another requirement for the LSG is for it to work with different people's voices and all languages. A further requirement is for the software code to be small enough to download from the session server 1 over a network 2 to the client personal computer 3 without too long a delay.

LSG approach and algorithms

The approach involved creating spectrograms of simple sounds and recording the corresponding facial geometry made whilst speaking those sounds. The spectrums in the spectrograms were then studied to look for patterns that could be transferred into heuristic algorithms. These algorithms were then installed in the jaw, mouth, lip and tongue analysers 123, 125, 127 and 129. Once the system was working for

simple sounds with algorithms in place in the analysers, it was tested with more complex words and different voices. Whenever the facial animation was found to be unacceptable the algorithms were adjusted or new algorithms developed to improve the facial animation.

In this embodiment, simple algorithms are disclosed for the analysers that work to an acceptable level on a variety of voices, languages and adequately for some singers. It is appreciated that these algorithms can be improved upon and this is a target of future research work.

The algorithm in the jaw rotation analyser 123 relates the output jaw angle to the energy in the high frequency bins. In general, whilst talking, the mouth opens further when making high frequency sounds than low frequency sounds. In the jaw rotation analyser 123, the higher the amplitude in the high frequency bins, the larger the jaw rotation and the more the mouth is open. The algorithm in the jaw rotation analyser 123 calculates a normalised average value 124 of the sum of the normalised amplitudes in the high frequency ranges f5, f6 and f7. This algorithm in the jaw rotation analyser 123 can be improved by setting a minimum level of mean normalised amplitude in the high frequency ranges f5, f6 and f7. If the actual mean normalised amplitude is not above this minimum level then the output value 124 is set to zero. This stops the mouth opening in response to low levels of background noise rather than speech.

The algorithm in the mouth length analyser 125 works on frequency range. The wider the range of frequencies, the larger the length between the mouth corners 126. The standard deviation of the spectrum is calculated from the amplitudes in each bin in the spectrum. The mouth length 126 output by the mouth length analyser 125 is proportional to this standard deviation. The mouth length 126 is a normalised value from 0 to 1. Whistling is an extreme example in which the mouth length 126 is very short to make a small hole through which air is expelled at a focused frequency. The mouth length analyser 125 can handle whistling because the standard deviation of a whistling sound is very small and the output mouth length 126 is correspondingly small.

The lip rotation analyser 127 looks for high amplitudes at particular frequencies. Lip rotations are associated with plosive sounds such as 's' or 't' that are in effect sudden bursts of energy at characteristic frequency range. Each plosive sound has a characteristic frequency bin or set of neighbouring bins. The higher the relative amplitude at one characteristic frequency, the larger the lip rotation. The lip rotation analyser 127 checks for high amplitude at one of these known sets of frequency bins relative to all the other frequency bins. The lip rotation 128 output by the lip rotation analyser 127 is proportional to the ratio between the average amplitude of the set of characteristic bins and the average amplitude of all the other frequency bins. The lip rotation 128 is a normalised value from 0 to 1.

The tongue protrusion analyser 129 looks for characteristic sounds such as 'th' in which the tongue protrudes. The higher the amplitude of the characteristic sound, the more the tongue protrudes.

Emotion detection

It is useful to detect the emotion of a person from the person's voice. Once detected, the emotion can be used to modify the avatar's actions such that the avatar's visual behaviour matches the emotion conveyed by the audio. Some emotions engender large changes in body language and other emotions engender barely noticeable changes in body language. For a good avatar metaphor it is useful to detect emotions that engender large changes in body language.

Referring again to Figure 28, the simplest emotion to detect is the absence of speech over time. This can be detected by a special emotion analyser 135 designed to detect absence of speech that outputs a strength of speech 136. If the strength of speech 136 is zero, then there is no speech at that time t. If the strength of speech 136 is 1 then there is speech.

An emotion that engenders large body movement is laughing. Laughing has a characteristic pattern that can be detected from speech. There is a regular pattern of sounds at a frequency of around 3-4 Hz along the time axis in the spectrogram 145 and characteristic high amplitude

such as levels a4-a6 and a low frequency such as f5-f7 in the spectrum 146. Laughing can be detected by a special emotion analyser 135 designed to detect laughing. The strength of laughing 136 output is a normalised value in the range 0 to 1.

Anger can be detected by an increase in amplitude. This is not always reliable, because, for example, moving the microphone 12 closer to the mouth of the user 17 may result in a significant increase in amplitude.

It is a further purpose of this embodiment that emotions be detected from the audio signal of a person speaking in near real-time and that the detected emotions be used to modify the movements of the avatar representing that person.

Geometry damping

It was found that the raw streams of real numbers from all of the analysers 123, 125, 127, 129 and 135 were noisy. The values went through substantial fluctuation from one spectrum analysis to the next. This gave poor facial animation in which vibrations of the order of 30 Hz with large amplitudes were observed during lip synchronisation with speech. After experimentation with damping, it was found that the best results came from damping each geometric parameter stream independently. For a parameter stream P:

$$Pmt = rPt + (1-r)Pmt-1$$

Pmt - the modified value of the parameter P at time t
Pt - the raw value of the parameter P
r - the damping ratio
Pmt-1 - the modified value of parameter P at time t-1

The damping ratio used was r=0.75. It is likely that different methods of damping will be developed for different geometric parameters and that these may have different values for any damping ratios r.

Identifying the main speaker

In an audio conference call, there is no need to identify the main speaker. The voice channels 91 can just be mixed and the users 17 will sort out the situation if several people speak at once; if necessary a chairman will be appointed to determine the next speaker.

In an avatar conference, it is useful to know who the main speaker is for several reasons:

- To plan camera shots
- To stop lip synchronisation being generated from just background noise giving the visual effect of many avatars speaking all the time when they are not actually speaking

Microphones 12 often pick up background noise 271, particularly if the user 17 is in an open plan office. In an ideal world, all microphones would only pick up the voice of the user 270 and automatically filter out background noise 271. In many user environmental locations 273, background noise 271 can be at the same amplitude as voice 270 or even higher. The filter 205 in the LSG 100 plays an important role in reducing this background noise 271 before it reaches the LSG 100. Where the background noise 271 is high, it is difficult for the LSG 100 to know whether the audio stream is noise 271 or voice 270, even after filtering. In many cases, the LSG 100 generates a stream of geometric positions 101 from the digital audio stream 104 that is in fact just background noise 271.

One simple way of eliminating the problem of identifying whether the audio stream 104 is voice 270 or background noise 271, is to request users 17 to turn off their microphones 12 when they are not speaking. This effect can also be implemented in a different way by requesting that the user 17 presses a 'Push to Speak' button 272 on the avatar session user interface 10 whilst he speaks. If several users 17 have their buttons 272 depressed at the same time, then the audio mixer 90 mixes all the active channels.

Users 17 are often multi-tasking with their hands and many users do not want to press a button each time they wish to speak. The ideal way of eliminating the problem of identifying whether the audio stream

104 is voice 270 or background noise 271, is to improve the filtering. Improved filtering will remove the need to press a button or switch the microphone on or ask the user to work in a quiet room. Filtering may be improved by using active noise reduction in which a second microphone situated away from the speaker's mouth can capture background noise and subtract it from the signal in the main microphone. As the power of personal computers grows, it will become possible to train software with the voice pattern of the speaker and to use that pattern to isolate the voice from the background noise.

It is a further purpose of this embodiment that the LSG 100 uses filtering and switching techniques as described above to more reliably generate events 81 that indicate whether a user 17 is speaking or not.

LSG Architecture

The avatar conference is a client-server architecture. The LSG 100 runs on each personal computer client 3. The alternative was to run the LSG 100 on the session server 1. It is better in most instances to run the LSG 100 on the personal computer client 3 rather than the session server 1 because (a) this uses up less network bandwidth in that the data rate for the combined compressed audio and geometric values 134 is much less than that for the digital audio stream 104 and (b) the network architecture is more scalable for large conferences in that massive session server processing demands are avoided.

The software code size for the LSG is around 20 kBytes. This has the advantage of being small compared to other approaches which often involve the necessity for large dictionaries to be on the client personal computer 3, usually by downloading over the network 2 from the session server 1. Such a small size of software code makes the LSG suitable for applications on small network devices such as mobile phones.

It is a purpose of this first embodiment to disclose a process wherein sound passes through the avatar user interface system comprising the following steps:

- a microphone means records sound from a user of a computing appliance means as the user speaks;

- a lip synchronisation generator means on the computing appliance means processes the sound to provide a combined audio and geometric position stream;
- the computing appliance means streams the combined audio and geometric position stream over the network to an audio mixer;
- the audio mixer mixes the combined audio and geometric position stream with any other combined audio and geometric position streams to produce a specific mixed audio and geometric position stream for each computing appliance;
- the audio mixer sends each computing appliance its specific mixed audio and geometric position stream;
- the computing appliance plays the specific mixed audio and geometric position stream to its user via a loudspeaker means.

It is a further purpose of this first embodiment to disclose a lip synchronisation generator process comprising a process performed at regular intervals on a digital audio stream flowing into a buffer of the following steps:

- the contents of the buffer are copied and then the buffer is emptied;
- a discrete fourier transform is performed on the copied contents of the buffer and a spectrum is output;
- one or more analysers analyse the output spectrum and each analyser outputs a value representing a geometric position of a part of a talking head.

Camera shot direction

One role of the software director 80, a software agent, is to decide and activate the cameras to form a sequence of shots. The camera shot shown in the meeting room media window depends on:

- the mode chosen 84
- the layout chosen 85
- flow of events (historical and actual) 81
- flow of shots (historical and actual) 82
- flow of actions (historical and actual) 83
- timers 86
- random choice

- the cameras programmed (61-64, 71-77 etc)

The rules for the shots can be very simple for some modes such as Mode M1 and fairly complex for modes such as M2. The person programming these rules has a large degree of freedom and is in effect building an expert system of an expert film director. The rules are improved with feedback from users during trials.

During the avatar conference it is normal for different people to speak at different times. Since each person has his own microphone 12 and personal computer 3, it is known which avatar is associated with a voice stream. Events include a person stopping speaking and another person starting speaking. The camera shot is usually on the main speaker; if several people are speaking at once then a wide shot of all the participants can be shown.

Complex shots that are difficult to do in the real world can be achieved with relative ease in software in a virtual world. Hollywood films are incorporating more and more shots filmed using a motorised robot arm to move the camera long distance in six degrees of freedom. This gives a 3D effect from the parallax of objects with verticals and horizontals moving. It has been found that this 3D effect increases the sense of presence in the viewer and enhances the enjoyment of the film. As an example of the use of this technique in an avatar conference, a moving camera can track a person as he enters the room and sits down. It is a further object of this invention to maximise the sense of presence for the viewer by using 6 degree of freedom camera movements.

Acting direction

Another role of the software director 80, a software agent, is to decide on the ambient and event animations of the avatars. This is equivalent to the director of a stage play defining every aspect of an actor's facial and body movement. The animation shown in the meeting room media window depends on at least some of:

- the mode chosen 84
- the layout chosen 85
- flow of events (historical and actual) 81

- flow of actions (historical and actual) 83
- timers 86
- random choice
- actions 83 available in a library 87
- action generator capabilities 88 as defined by action parameters 243

Animation actions

Animation actions can be classified into four types:

- Ambient animations (generated by software director)
- Event animations (generated by software director)
- Head/facial animated gestures (triggered by user)
- Hand/arm/body animated gestures (triggered by user)

Ambient animations

An actual person is almost never still. Breathing, swaying, changing gaze, small head movements and many others are termed ambient animations. In a meeting, ambient animations depend on the role of the person and his culture. A speaker will usually move his hands and arms a lot. A listener will be less dynamic. Ambient animations are designed to be encouraging towards a good meeting atmosphere; listener's faces can be seen to smile and look positive; heads can nod regularly as if in agreement or in understanding; body posture can be upright rather than slouched. Ambient animations are generated automatically by the software director.

Event animations

Event animations are the actions associated with an event. Here are some examples:

- a person entering the meeting room, walking to his chair, pulling the chair out, sitting on it and moving the chair nearer to the table
- the detection of emotion from the audio stream; for example, if a laugh is detected, the avatar can be animated as laughing
- a participant has been silent for longer than a certain period, actions associated with the participant not being involved in the meeting are adopted; a method might be a certain slouching in the

chair that will convey visually to the other participants that this person is not involved much

- the participant is not able to see the meeting room media window 50 because he is viewing another document, then his avatar could be seen reading a document
- a participant takes another session (call). His avatar can be seen using a mobile phone

Event animations are generated automatically by the software director in response to an event.

It is a further purpose of this embodiment that the software director automatically generates ambient and event animations.

Gaze

Humans have clear and distinctive patterns of gaze when engaged in face to face situations. If the software director creates patterns of gaze between the avatars during the conference that meet the subconscious expectations of the viewer, then the viewer will experience a high sense of presence. If the software director creates patterns of gaze between the avatars that break the subconscious expectations of the viewer, then the viewer will be distracted and find the avatar behaviour to be disconcerting. It is a generally accepted research conclusion, that one of the limiting factors on the uptake of video telephony is that the patterns of gaze are disconcerting. The software director 80 uses rules for controlling the gaze of the avatars based on observations of people in meetings.

Gestures

In a meeting, a participant often wishes to convey information by body language gestures. The gesture is sometimes purposeful - based on an active decision by the participant. Examples include:

- raising his hand to show he wants to ask a question
- clapping in applause
- waving to say hello

Body language can also be passive, often without the participant being aware of the body language he is sending out. Examples include:

- shaking head in disagreement with what is being said
- nodding in agreement with what is being said
- slumped in a chair, bored

In an avatar conference, a participant could select a button in the user interface corresponding to the body language gesture he wishes to convey. Other participants looking at the meeting room media window will see the gesture. Both active and passive gestures could be used. Gestures can be particularly useful to the chairman of a meeting who can respond to a gesture in choosing the next person to speak.

It is a further purpose of this embodiment that the software director generates animated gestures in response to an active user trigger.

Animation architecture

The software director 80 generates a flow of animations 83 for each avatar 5. The animations are retrieved from an action library 87 or are generated in real time from an action generator 88.

Actions 83 in the action library 87 are fixed actions with a fixed duration and fixed movement. They are usually created by motion capture or by key frame animation. An example is raising a hand to wave.

Actions 83 generated by the action generator 88 are variable actions that are generated in real-time to action parameters 243 specified by the software director 80. An example is asking the action generator 88 to generate a walking animation action 83 that follows a specified path across the meeting room floor. A possible set of action parameters 243 for this example are: the avatar number 8, the path specification, the walking style, the speed, starting conditions and end conditions.

If a meeting room is designed with known dimensions, then an action library 87 of all possible actions 83 can be compiled from motion

capture of an actor or key-frame animation. In this case, an action generator 88 is not used.

It is a further purpose of this embodiment that any action 83 for an avatar 5 during the conference can be chosen by the software director 80 either from an action library 87 or an action generator 88.

Animation blending

Often the 3D position of the avatar at the end of one action 83 is not directly compatible with the 3D position at the beginning of the next action 83. The result is a 'jump' from one frame to another in which hands or feet may travel as much as a metre over 1/25 second, or whatever the time is between frames. This is very unrealistic and annoying to the user. Blending of joint positions over several frames is used to reduce this problem.

In an animation, the movement of an avatar can be defined as a set of joint positions at each time point or frame in the animation. The positions of each vertex on the skin or clothes of the avatar are determined from the joint positions and any weightings associating a vertex with each neighbouring joint. The main advantage of defining an animation as a series of sets of joint positions is that it is smaller than a series of sets of vertex positions. An avatar typically has 20-50 joints but thousands of vertices. A file with a set of joint positions stored for every 1/25 second is many times smaller than a similar file with vertex positions. To blend two actions 83 it is necessary to adjust a number of frames of animation in the first action 83 prior to the join and to adjust a number of frames of animation in the second action 83 after the join such that the last set of adjusted joint positions in the first action is geometrically very similar to the first set of adjusted joint positions in the second action.

Blending is pretty good for joining two similar positions: this is known as a subtle blend. However, when the positions are radically different, the result can be completely incorrect. It is quite possible for arms and legs to pass through each other during a radical blend; this effect can be most annoying for the user. The software

designer designing an avatar conference system must carefully define each action 83 in the library of actions 87 such that all possible actions 83 that the software director 80 selects to follow any given action 83 require only a subtle blend and not a radical blend. The main method used to achieve this is the adoption of a limited number of neutral positions, with each action 83 edited until it starts in one neutral position and stops in another neutral position.

Animation merging

More than one action 83 can be merged and played simultaneously to form a single merged action. Actions are one of two types:

- Dominant action
- Modifying action

A dominant action is an action involving major displacements such as walking. A modifying action is an action involving minor displacements such as ambient actions and smiling. Each action 83 in the library has a defined action type: either a dominant action or a modifying action. The most common modifying action is facial animation. It is possible to merge three or more actions. But only one action in a merged action can be a dominant action. For instance, the walking dominant action can be defined with smiling and lip synchronisation. Modifying actions are applied to the dominant action one frame at a time. The modifying action is defined as a relative movement of joints. A modifying action is 'added' on top of a dominant action during animation.

Animation re-targeting problem

Each avatar 5 of a particular person is a unique size. Some avatars may be short and fat, others may be tall and thin. When an action 83 is created for the action library 87 it is created on an avatar of a particular size. If the creation means is motion capture, then the action 83 will play back best on an avatar with the same size and shape as the person whose motion is captured. Similarly, if a skilled animator creates an action 83 for an avatar of a particular size, it will play back best on an avatar with similar size and shape.

The use of joint positions to define animations, makes it possible for animations created on an avatar of a particular size, to be played back on avatars of different sizes. It is a further purpose of this embodiment that any action 83 can be played on an avatar 5 of a different size and shape from the avatar 5 for which the action 83 was created.

Problems occur when there is interaction between an avatar and attributes of the virtual environment such as chairs, tables, floor, door handles and cups. Replaying an action 83 involving contact with an attribute of the virtual environment on any size avatar may result in poor motion artefacts. Examples of poor motion artefacts include: avatar arms passing through tables, not grasping cups properly and hovering above chairs.

Re-targeting solutions

This problem may be overcome with a commercially reasonable amount of effort by the simplifications of:

- morphing all avatars 5 to the same standard size and shape
- preparing all actions 83 for avatars of that standard size and shape in a defined virtual environment
- crafting the software director 80 state machine to generate series of actions that work without exhibiting poor motion artefacts

However, the photo-realism of the avatars will be severely degraded, if the avatars are all the same size and shape.

Different application activities present different re-targeting solutions:

(i) Camera control: The camera 72 viewpoint, direction, zoom may be controlled by the software director 80 such that poor motion artefacts are not shown to the viewer.

(ii) Aspect ratio: In a meeting activity, the avatar user interface window 260 aspect ratio might be high with the window wide and thin such that only the upper bodies of the avatars are visible; in this way, accurate animation of the feet and posterior is not needed.

(iii) Avatar size range: Avatars are scaled across a design size range between a minimum and a maximum. Very small avatars are

scaled up to a minimum size, very large avatars are scaled down to a maximum size and the rest are spread between. In this way, taller people have taller avatars than shorter people's avatars. The environment is designed to cope with avatars in the design size range. The software director 80 state machine is crafted to generate series of actions that work without exhibiting poor motion artefacts for avatars within the design size range.

(iv) Adaptive action control: Actions 83 such as sitting in a chair may be animated adaptively to avoid particular motion artefacts. An example is sitting in a chair. Avatars of different heights and different posterior sizes might either float above the chair seat or break through it. Adapting the sitting down action 83 to the avatar size by raising or lowering the whole avatar during the sitting process, solves this problem. In this case, the action 83 is probably generated by the action generator 88 based on action parameters 243.

(v) Morphed body parts: For example, to help with the grasping problem, all avatars could be given the same size arms and hands. In this way, it is only necessary to position the avatar's shoulder joint in a fixed position relative to the prop 383 for the action 83 to be executed without poor motion artefacts.

It is a further purpose of this embodiment that one or more re-targeting solutions are used to avoid poor motion artefacts.

Speech and text means

According to this embodiment, Figure 31 is a block diagram of the session server 1 containing: an audio recording 185, an event accumulator 89, a speech recognition engine 182, voice profiles of participants 184, a text transcript 183, a translation engine 186, translated text 187, a text to speech engine 188, a voice profile 184 of the voice used in the text to speech engine 188, a text chat engine 189 and an e-mail engine 190. The software engines are running in memory 346. The session server 1 is connected over a network 2 to a speech recognition service 192, a text to speech service 193 and a

translation service 191.

Conference recording and playback

It is quite common for a person who should have been in the conference to miss it and to wish to know what happened. The conference can be recorded for later playback by the person who missed the conference.

The conference can be stored as a linear audio file 185 and a time-stamped event accumulator 89. Events include:

- person enters conference
- new speaker starts
- new agenda item started

On playback, the audio recording 185 can be compressed in length to reduce the amount of time a person needs to spend listening to the audio recording. For example, periods of time in which there was no speech can be removed. Also, the time axis can be compressed such that playback takes less time than the original conference took. The playback speed eg 125% of normal speed, can be controlled by the person listening. The person playing back the conference can also use the event accumulator 89 as key points at which to start listening to the recording. For example, if he is only interested in agenda item number 3 then he can skip to the point at which the chairman has noted that agenda item number three started.

Speech recognition

As speech recognition engine technology improves, it may become feasible for a high enough quality text transcript 183 of the meeting that is acceptable to users to be automatically produced by a speech recognition engine 182 from the audio recording 185 and event accumulator 89 using voice profiles 184 to improve the speech recognition. The text transcript 183 can be generated by the speech recognition engine 182 after the conference or in near real-time during the conference. A speech recognition service 192 may be used instead of having a speech recognition engine 182 on the session server 1.

Text translation

The text transcript 183 can also be translated to translated text 187 in another language using a translation engine 186 present on the session server 1 or by a network translation service 191 over a network 2. The translated text 187 can be generated from the text transcript 183 after the conference or in near real-time during the conference.

Text to speech: audio translation

As text to speech engine technology, text translation engine technology and speech recognition engine technology improve in quality and speed, it may become feasible for a high enough quality near real-time audio translation of the meeting that is acceptable to users to be automatically produced by a speech recognition engine 182, a text translation engine 186 and a text to speech engine 188 from the audio 104. Eventually, each participant can define the language spoken and the language to be listened to such that a true multi-lingual avatar conference can take place. A text to speech conversion service 193 may be used instead of having a text to speech engine 188 on the session server 1.

Text Chat

During the conference, participants can see text chat in a dedicated window 26 driven by a text chat engine 189 on the session server 1. A participant can input and send text messages to all participants or just to selected participants.

The text chat window 26 can be used to show any or all of: text sent by a text chat engine 189, events 89 described in a textual format, a text transcript 183 and translated text 187. The text chat window can be set to the preferred language of the user such that all text is translated and displayed in the text chat window 26 in the preferred language. Text can be shown twice: in the language in which it was generated and in translation.

e-mail

Following the conference, the e-mail engine 190 can send copies of some or all of the text generated during the conference in e-mail form

to the e-mail addresses of participants and also to those who could not attend.

The e-mail engine 190 can also be used as an e-mail reflector for participants in which e-mails concerning the conference whether before, during or after the conference, are sent to the e-mail engine which will then immediately forward copies to all participants.

Participant roles

During an avatar conference, users 17 can have identical roles from the point of view of system functionality or they can be assigned different roles with different avatar session user interfaces 10.

A user 17 in a Chairman role can be provided with functionality to enable him to:

- Remove a user from the conference
- Select a speaker to speak next

A user 17 in a Secretary role can type minutes.

A user 17 in a Teacher role can control the display seen on all personal computers 3 in a presentation.

Participant performance

During an avatar conference, the activity of users 17 can be recorded and fed back to participants. If a user 17 has not spoken for a period of time, an event animation is used such that his avatar 5 can be animated in a way that shows his lack of recent participation. The avatar might sink down in the chair and appear to withdraw from the conference. If this visual withdrawal is noticed by other participants, then they have the opportunity to try to involve the quiet user in the conference. Alternatively, statistics of % of the conference time that each person has spoken for might be shown. This will show up users who might be hogging the conversation and others who might be lurking without saying anything. Real-time performance feedback can provide the participants as a team with a tool for making their conferences more effective.

In applications such as education or training, participant performance data such as attendance records can be available to teachers. Storage and access to information on participants performance is liable to be regulated by laws in different countries.

Some of the performance data available includes:

- Number of times a person has spoken at the conference
- Average length of time of each interaction
- Total length of time speaking
- % of conference time speaking
- Attendance at a series of conferences: % of conferences
- Attendance at a series of conferences: % of time
- Number of slides presented
- Number of times whiteboard used
- Number of times chat used
- Number of times whispering used

Webcam use

There are occasions when it is useful to see live video 336 of a participant or an event at a location that a participant wishes the conference to see. Referring again to Figure 11, the video (or streaming webcast) 336 can come from a webcam 29 situated on the display device 264 of a user 17. Alternatively, the video 336 can come from any other type of video camera 29 connected to a personal computer 3 on the network 2. The quality of the streaming video 336 seen by each participant will vary with the bandwidth available to the participant. It can vary from one frame every few seconds for one participant with a low bandwidth connection to full frame rate for a participant with a high bandwidth network connection.

The resolution in pixels of the webcam broadcast 336 is usually small and the software director 80 shows the webcam in a correspondingly small window. To avoid seeing both a person live and his avatar at the same time, the avatar from whose webcam 29 the broadcast 336 is streaming must leave. To maintain the metaphor, the avatar walks out of the room before the webcast 336 starts and walks back in when it

finishes. The streaming video webcast 336 from the webcam 29 is shown on the screen 53.

SECOND EMBODIMENT

According to this second embodiment, Figure 32 is a block diagram of an apparatus for holding an avatar user interface session in accordance with a second embodiment of the present invention. In this embodiment, the apparatus comprises a plurality of personal computers 3 that are connected by a network 2 to a session server 1, an avatar hosting server 4 containing avatars 5 and a telephone network 155 with telephones 150 and a telephone server 154.

In this second embodiment voice is carried over either the telephone network 155 or the network 2 and data is carried over either the telephone network 155 or the network 2.

IP/PSTN audio architecture

Currently, there is large lag in the existing public internet and the quality of voice over the internet protocol (VoIP) is much less than for the PSTN telephone network or mobile networks such as GSM or 3G.

Two main protocols exist for transmitting over an IP network: HTTP and UDP. HTTP checks that each packet is received. This checking is the main cause of lag. UDP does not check and typically has much less lag. However, UDP is considered a security risk for companies and companies typically configure their firewalls to prevent UDP from getting through. A UDP system that does not work for most companies will not be purchased. In the future, new versions of IP such as IP v6, may improve the quality and access of VoIP such that it rivals that of telephone networks.

The main method for remote conferencing today is telephone conference calls using the PSTN, mobile networks and a conference server for mixing the calls. Telephone conference calls are expensive, not only for the calls but also for the service of the session server. Anyone can access these conferences from wired or wireless handsets.

For an avatar user interface session using VoIP, access is limited to those who have a microphone and headphone on their computers and who are situated by a networked computer. Someone who does not have a networked computer with microphone and headphone cannot participate in an avatar user interface session using VoIP as disclosed in the first embodiment.

The use of IP for audio, as disclosed in the first embodiment, avoids the cost of the telephone calls. To make an avatar user interface session convenient and available to all those wishing to participate, it can be an advantage to have combined IP and telephone networks as disclosed in this second embodiment.

According to this second embodiment, a telephone server 154 is connected to the IP network 2. Party#1 151 can use his telephone 150 over a telephone line 155 to a telephone server 154 and his personal computer 3 on network 2. Party#2 152 can use his headset 11 and his personal computer 3 over network 2. Party#3 153 can use his telephone 150 over a telephone line 155 to the telephone server 154 and not see the avatar user interface session visually. Party#4 158 can use his mobile telephone 157 over a mobile telephone network 159 to a mobile telephone server 156. When mobile handsets advance and 3G mobile infrastructure is in place it will be possible for audio and data to be used simultaneously on a mobile handset. In this way Party#4 can transfer both voice and data over the mobile network 159. Party#4 could wear a hands-free headset for audio and look at the screen of his mobile handset to see the avatar user interface session.

The audio mixer 90 can be resident on either the session server 1 or on a telephone server 154 or 156 on a separate computer.

The Lip Sync Generator (LSG) 100 is normally present on the personal computer 3 through which it is connected to the sound card 102. When a telephone connection 155 is used, the LSG functionality 100 can be present on a server, either the session server 1 or the telephone server 154. The geometric positions stream 101 and the audio transform stream 105 can then be routed to the personal computers 3 over the network 2 or to a mobile device over the mobile network 159.

It is a further purpose of this second embodiment that voice or data transfers in the avatar user interface session can be over a plurality of networks of any types connected by devices such as network switches or network routers. Examples of networks include: the internet, intranets, extranets, Virtual Private Networks (VPNs), GSM mobile networks, GPRS mobile networks, 3G mobile networks, satellite networks.

It is a further purpose of this second embodiment that communication appliances in the avatar user interface session can be any sorts of devices including but not limited to: personal computers, mobile telephones, networked personal digital assistants, networked computer games consoles, interactive digital televisions, laptop computers.

It is a further purpose of this second embodiment that the system architecture can be of any type including client server and peer to peer and that any item of system functionality disclosed in this embodiment can be resident on any device. Any communication appliance might also act as a server as well as a client. As an example, it will be appreciated that the session server 1 does not need to be an independent unit and that a computing appliance 3 could run both the functionality of the session server 1 and the avatar user interface 160. It will similarly be appreciated that the software functionalities and hardware capabilities of many servers could be combined into a single computing appliance 3. As a further example, the session server 1, the avatar hosting server 4, the avatar hosting registry 226 and the avatar agent hosting server 321 could be combined in one computing appliance 3.

THIRD EMBODIMENT

In this third embodiment the format of the avatars in each communication appliance is appropriate to the computing power, graphics processing power and display size of the computing appliance such that real-time visualisation in the avatar user interface system can be achieved.

Animations of avatars 5 at much less than 12 frames per second look

jerky and reduce the sense of presence felt during a session on an avatar user interface system.

Different 3D representations

Avatar computer models can be in different mathematical 3D representations. Possible representations include but are not limited to: triangles, quadrangles, other n-sided polygons, B-spline surfaces, NURBS and subdivision surfaces. It is a further purpose of this third embodiment that the format of a 3D avatar 39 can be any 3D mathematical representation.

Progressive 3D representation

Some representations can be progressive 3D representations in which an actual format displayed can be an instantiation of a representation of arbitrary size on a continuum from low size to high size. In this way, an instantiation can be chosen that is optimal for the power of the computing appliance.

Animatable image representation

In addition to 3D representations, avatars may be represented in other ways. One way includes an animated image representation.

According to this third embodiment, Figure 33 is a schematic diagram of an animatable image 380. There are a minimum of two parts to the image: an animatable image avatar 382 in the foreground and a background image 381.

In a basic representation, an animatable image 380 may be described as a talking post card in which a talking avatar 382 and optionally a prop image 383 are superimposed in front of a fixed background image 381. The background image 381 is usually photo-realistic. The animatable image avatar 382 is usually photo-realistic.

According to this third embodiment, Figure 34 is a schematic diagram of an animatable image avatar 382. For the purposes of animation, the animatable image avatar 382 is considered to be split into five animatable avatar segments 395:

(i) upper body segment 390

(ii) jaw and mouth segment 391

(iii) eyes and eyebrows segment 392

(iv) head segment 393

(v) face segment 394

Each animatable avatar segment 395 has a set of one or more different images representing that segment. The upper body 390 segment normally has only one image in its set.

Figure 35 is a schematic diagram of a set of four state images 425 for the jaw and mouth segment 391 showing the jaw and mouth in four states: neutral 470, happy 471, sad 472 and laughing 473. It is usual, for a high fidelity animation to be possible, that the jaw and mouth segment 391 has several more state images in its set 425. The eyes and eyebrows segment 392 has at least two state images 425: eyes closed and eyes open. The head segment 393 normally has several state images 425 in its set with the head at slightly different orientations. The face segment 394 normally has several state images 425 for different facial expressions in which wrinkles play an important role.

Figure 36 is a tree diagram of the hierarchy of animatable avatar image components. A complete set of images 424 for playing an animatable image 380 comprises the background 381, prop 383 and for each avatar segment 395 the set of state images 425.

The animatable image avatar segments 395 in this embodiment are not limited to the five disclosed animatable image avatar segments 395. The animatable image avatar 382 might be split into more or less segments.

Animatable image generation

Figure 37 is a schematic diagram of an animatable image generator 397 resident on an avatar hosting server 4. The animatable image generator 397 is based on an avatar player engine 210. An animatable image 380 comprising a complete set of images 424 may be generated by the avatar player engine 210 from a photo-realistic avatar 238 and a virtual background scene 65 using a virtual camera 61. The photo-

realistic avatar 238 is posed in front of camera 61 to form a neutral pose defined as an action 83 in which the photo-realistic avatar 238 looks forward, eyes open, neutral expression and mouth closed. This pose when viewed from camera 61 generates a base animatable image avatar 382. When the photo-realistic avatar 238 is removed, the image of the scene 65 viewed from camera 61 is the background image 381. The set of state images 425 for an animatable avatar segment 395 are generated by applying a predefined set of poses as actions 83 in the animatable image generator 397.

Figure 38 is a schematic diagram of an apparatus for animatable image generation 398. If a photo-realistic avatar 238 of a subject person 428 is not available, an animatable image 380 may be generated from a single photo-realistic image 399 of a person in front of a background. A skilled person 427 will use image processing software 426 running in memory 345 on a personal computer 3 to process the image 399 to define the complete set of images 424.

This invention is not limited to using animatable image generators 397 and 398. For example, a complete set of images 424 could be generated from video 336 or a set of still images taken of the subject person. The animatable image generator 397 could be resident on a personal computer 3.

Animatable image playing

Referring again to Figure 20, the animation of an animatable image 380 is generated and played in a similar way to that of a 3D avatar 39. A software director 80 generates the actions 83 that are played by the player 210. The main difference is that the software director 80, the player 210 and the other components in Figure 20 are designed to work with animatable images 380 instead of 3D avatars 39 and scenes.

The animatable image avatar 382 is animated in the player 210 from actions 83 by a combination of methods that are now disclosed. The animatable image avatar 382 is normally based on a front view of the avatar covering at least the face, but rarely descending below the shoulders. This focus on the face removes the need to attempt to animate upper body movements such as arm gestures and lower body

movements such as walking or even turning the head by more than a few degrees.

There are five animation action types 83 generated by the software director 80 for the animation of an animatable image avatar 382 by a player 210:

A. Body movement

B. Head movement

C. Lip synchronisation

D. Eye movement

E. Facial expression

The body movement action A is limited to a combination of horizontal translation, vertical translation and rotation relative to the background image 381. A body movement action A. affects all five animatable avatar segments 390-394. The five animatable segments 390-394 are moved according to a body movement action A as if they were locked together.

The head movement action B is limited to two rotational components about the middle of the neck. A first rotation component left-right equivalent to shaking ones head and a second rotation component up-down equivalent to nodding ones head. A head movement action B affects the four animatable avatar segments 391-394. The four animatable segments 391-394 are moved according to a head movement action B as if they were locked together.

The two actions A and B are added together to give a combined head and body movement.

The lip synchronisation movement action C affects only the jaw and mouth segment 391. The eye movement action D affects only the eye and eyebrow segment 392. The facial expression action E affects only the facial segment 394.

The three actions C, D and E are applied locally to their respective segments after the actions A and B have been applied.

Two forms of morphing are used: at segment boundaries and between images in a set. At segment boundaries such as the neck which lies between the body segment 390 and the head segment 393, image morphing is used to stretch the image on one or both sides of the boundary. Between image morphing is used where there is a gradual progression from one image in the set to another for a particular segment.

Animation of image avatars 382 is not limited to the apparatus and methods disclosed above, but may be extended to any image based method.

This embodiment is not limited to one animatable image avatar 382 superimposed in front of the background image 381 but may contain two or more animatable image avatars 382. Referring again to Figures 16b, 16c and 16d, it can be seen that several animatable image avatars 382 may be generated from several avatars 5.

Referring again to Figures 17a, 17b and 17c, three layouts Layout 1, Layout 2 and Layout 3 may be used for displaying multiple animatable image avatars 382 with one or more background images 381 on a single display device 264. This invention is not limited to displaying Layouts 1-3 but may cover any layout that fits the application that this avatar user interface system invention is used for.

Props 215 may be converted into prop images 383 which are animated in front of the background image 381. The animated prop image 383 may appear as part of a background image 381; an example is a tree bending in the wind. Or the animated prop image 383 may appear separate from the background image 381; an example is a bird flying across the background image.

It is a further purpose of this third embodiment that the format of an avatar 5 can be any animatable non-3D mathematical representation including animated image representations.

Computing appliance variety

A computing appliance may be very powerful with a processor running at speeds in excess of 2 GHz, more than 512 MB of memory 345, a display

device 264 with more than 1 million pixels and a specialist 3D graphics chip such as an Nvidia GeForce 3 from Nvidia Inc (USA). Such a computing appliance can easily render real-time animation at 20 frames per second of 10-20 avatars 5 in the full generic format as disclosed in the first embodiment.

However, many computing appliances are less powerful and do not have specialist 3D processing hardware. Processing power is usually constrained so as not to use up battery life on lightweight portable devices with small batteries. Examples include mobile phones and wireless personal digital assistant appliances. Less powerful computing appliances usually have less memory than more powerful computing appliances. Less powerful computing appliances such as mobile phones may have very small display device 264 sizes with fewer than 5,000 pixels.

3D avatars with lower levels of detail may be used on intermediate power computing appliances to achieve the desired animation performance. Avatars with lower levels of detail typically have fewer polygons and smaller texture maps. This is good for achieving higher frame rates and uses less memory but the downside is that the visual quality of the 3D avatar is less good.

For high power computing appliances with a lot of avatars in the scene, a combination of low and high levels of detail avatars may be used to achieve a good frame rate. The closest avatars to the camera might be high level of detail and those furthest away might be low level of detail avatars. This can be achieved by having two or more level of detail avatars available and switching between them. Alternatively a progressive avatar approach might be used.

Animated image representations may be used on low power computing appliances to achieve the desired animation performance. These use less computing power and memory than 3D representations.

Multiple avatar formats and converters

According to this third embodiment, Figure 39 is a block diagram of an apparatus for holding an avatar user interface session in accordance

with a third embodiment of the present invention. In this third embodiment, the apparatus comprises computing appliance 160 with a specific avatar 5 in format A1, computing appliance 161 with a specific avatar 5 in format A2, computing appliance 167 with a specific avatar 5 in format A3 and an avatar converter software 164 of type C3 stored in memory 345. The computing appliances 162, 163 and 167 are connected by a network 2 to an avatar hosting server 4 containing a substantial number of avatars 5, database 6, avatar converter software 164 of types C1, C2 stored in memory 344 and specific avatars 5 of formats A1 and A2.

It is a purpose of this third embodiment that the avatar hosting server 4 has avatar converter software 164 such as C1 that can convert an avatar 5 into a specific avatar 5 with a format such as A1 at a different level of detail. The specific avatar 5 in format A1 is then transmitted over the network 2 to a computing appliance 160 for which the specific avatar 5 of format A1 is suitable.

An alternative approach is to have avatar converter software 164 C3 in memory 345 on a computing appliance 167 such that an avatar 5 can be converted to a specific avatar 5 format A3 locally on the computing appliance 167. Software techniques such as progressive meshes or variable levels of detail employed in the avatar converter software 164 C3 known to those skilled in the art might convert the avatar 5 to several different formats during the conference depending on the graphics load on the computing appliance.

It is a further purpose of this third embodiment that a computing appliance 167 can contain avatar converter software 164 C3 for which the specific avatar 5 of format A3 is suitable at any one instant. This invention is not limited to one type of avatar converter software 164 running on a computing appliance 167 but may allow any number of avatar converter software 164 on a computing appliance 167.

Different avatars 5 will contain different visual data depending on how they were originally generated; for instance, photo-realistic avatars 238, parameter avatars 232 and animatable image avatars 382 will be based on different raw data.

The avatar hosting server 4 usually stores the raw data from which the avatar 5 was generated, including manual input used to generate the avatar 5. The raw data for photo-realistic avatars is usually in the form of digital images 19. In this way, an avatar 5 can be regenerated automatically from the images 19. Moreover, if any technological improvements are made to the avatar 5, then a newer version of the avatar 5 can be generated automatically from the images 19 and replace the older version.

The suite of avatar converter software 164 should ideally be capable of converting any avatar 5 to any requested format. These conversions will not always be of the highest quality due to missing information. For instance, an animatable image avatar 382 cannot easily be converted into a photo-realistic avatar 238 because there is no information on the body shape.

The avatar hosting server 4 usually stores all formats of the avatar 5 that have been previously requested. The reason is to maximise response time for requests for that avatar in a particular format. If an avatar must be first converted into a particular format then it will take the avatar hosting server 4 longer to service a request. It is a benefit to the user that his request for an avatar is serviced as quickly as possible. However, storing several formats of each avatar uses up a lot of server space. To conserve server storage space, it may be pragmatic data management to delete formats that have not been used for a considerable time and formats that have been superseded by new versions.

It is a further purpose of this third embodiment that the communication session on the avatar user interface system invention may involve any combination of 3D or animatable image representations on the computing appliances. At one extreme, all the computing appliances may be high power personal computers 3 and use photo-realistic avatar 238 representations. At the other extreme, all the computing appliances may be mobile phones and use animatable image 380 representations. In a typical session, one computing appliance might use 3D avatars 39 with high numbers of polygons, a second computing

appliance might use NURBS based 3D avatars 39 and a third computing appliance might use animatable images 380.

Major conference

In a major conference, with around 50-1,000 participants, a number of techniques are used to run the avatar user interface session on any computing appliance 167. It is likely to be impossible for some time that a personal computer 3 would have enough power to fully animate and render 1,000 avatars 5 at the same time. In large conferences, it is still useful for the complete audience to be seen to provide participants with an ambience matching the scale of the event. There are many ways for the software director 80 to achieve this ambience:

- pan quickly across a single image taken at a real conference of the appropriate size; the real conference room must match the avatar user interface session room
- store short video clips in the personal computer 3 taken at a real conference of the appropriate size and replay them from time to time
- when there is a question from the audience, zoom in quickly from the real audience image to a virtual close up of the avatar surrounded by other avatars

The chairman of a large avatar user interface session needs to handle questions from a lot of people.

Figure 40 is a schematic layout of a major conference user interface functionality 291 for the Chairman consisting of a list of attendees with names 244 and organisations 293 wishing to ask questions, a button 294 for the Chairman to permit an attendee to speak. Attendees have buttons 290 to indicate a desire to ask a question and buttons 295 for testing their microphones before asking a question.

The avatar 5 of a user attendee 17 who has pushed the ask button 290, the software director 80 will raise the hand of the avatar 5. When the Chairman presses the button 294 to give the attendee 17 permission to speak, the software director 80 will lower the hand of the avatar 5 and connect the user 17's audio input channel 91 to the audio mixer 90.

On pushing the Test Microphone button 295, a user 17's microphone 12 will be connected over the network 12 to the audio mixer 9. A short dialogue will take place using pre-recorded sound files on the audio mixer in which the user 17 is able to verify that his microphone 12 works and is connected to the audio mixer 90. This test procedure should reduce the frequency of occurrence of a user 17 trying to speak but not being heard by the conference attendees because of a microphone problem.

It is often the case in large conferences that there are breaks between presentations during which people chat. The whispering capability of this invention will permit a large number of whispered conversations of 2 or more people during these breaks.

It is a purpose of this invention that large conferences of many thousands of people can be successfully held.

It is a purpose of this third embodiment to disclose a process wherein a remote presenting user presents a presentation remotely comprising the following steps:

- the remote presenting user starts a prepared presentation;
- remote audience users watch the avatar of the remote presenting user perform the prepared presentation;
- present audience users present physically together in a theatre watch a projection of the avatar of the remote presenting user perform the prepared presentation;
- the prepared presentation ends;
- a remote audience user asks a question;
- the remote presenting user views the avatar of the remote audience user asking the question from amongst a single virtual audience and the avatar of the remote audience user gazes at the remote presenting user;
- the present audience users view the avatar of the remote audience user asking the question from amongst a single virtual audience around the avatar of the remote presenting user and the avatar of

the remote audience user gazes at the avatar of the remote presenting user.

FOURTH EMBODIMENT

In this fourth embodiment, rather than each participant being in a separate location, the apparatus of the invention supports two or more participants at one location.

Speaker phone

It is common in audio conferences for several people to congregate around a speaker phone in a single room for a conference call which includes at least one other location. Often the speaker phone has several microphones attached to it that are placed near different people around the meeting table. In this way, the people in the room can communicate directly via physical document exchange, body language, whispering and facial expressions in parallel to the formal audio exchanges.

Shared display device

Figure 41 is a schematic layout of an apparatus for holding an avatar user interface session in accordance with a fourth embodiment of the present invention. A personal computer 3 with a computer cabinet 16 contains a wireless transmitter/receiver 170. Participants 17 'Albert', 'Bruce' and 'Charles' sit around a table 172 at an environmental location 273 with each participant 17 wearing a wireless headset 171 including microphone 12 and earphone 13. Each wireless headset 171 has an identified owner eg Albert. A large display device 264 shows the avatars 5 of all participants on the avatar user interface session other than those participants 17 around the table 172 at this location. Means for controlling the computer such as a keyboard 14 or a mouse 15 are available for use by the participants 17. The environmental location 273 is usually a room such as a meeting room.

In this way a participant 17 eg Albert can see all other participants either physically 17 ie Bruce and Charles, or as avatars 5. When a participant speaks, since the wireless headset 171 is identified as being owned by a specified person eg Albert, it is possible for the

lipsync to be applied to the correct avatar 5 of Albert. The wireless headset may be identified by means of an identification chip inside the headset.

Sound mixing

In a further refinement of this embodiment, one or more loudspeakers 173 are used for broadcasting sound to the participants 17 and each user has a wireless microphone 12 linked to the identity of the user. Signals from the wireless microphone 12 are transmitted to the receiver 170. To prevent audio feedback between the loudspeakers 173 and the microphones 12, the audio mixer 90 does not mix in the audio streams from the microphones 12 of all the participants at that location.

In the case where there are two or more loudspeakers 173, 3D sound can be used to increase the sense of co-presence of the participants. If an avatar 5 on the far left of the display device 264 is talking, then the 3D sound can be mixed locally to appear as if it is coming from the mouth of that avatar. In this case, the sound volume from a loudspeaker on the left would be louder than that from a loudspeaker on the right. The audio mixer 90 is not involved in generating the 3D sound.

Figure 41a is a schematic of the 3D sound processing. The mixer 90 generates an audio output stream 93 which travels over the network 2 to the PC 3. A splitter 141 splits off the geometric positions 101 to the player 210. The splitter 141 sends the remaining audio transform 105 to the decompressor 140. The decompressor 140 generates digital voice 104 and streams it to the 3D sound generator 143. The player 20 calculates the pixel coordinates 142 on the display 264 of the mouth of the avatar 5 that is speaking and streams them to the 3D sound generator 143. The 3D sound generator 143, uses the known positions of the loudspeakers 173 relative to the display 264, to generate digital voice signals 104 to the sound card 102 which streams analogue voice 103 to the loudspeakers 173.

This fourth embodiment has the advantage of allowing an avatar user interface session to take place with more than one person at a single location. It is also scalable for the case where there are two or

more locations, with more than one participant at each location. Furthermore, it has the advantage of greatly increasing the sense of presence by showing all the non-present people on the call as avatars.

FIFTH EMBODIMENT

It is a purpose of this fifth embodiment, that the avatar user interface system comprises an integrated multi-media communication system based around photo-realistic avatars for communication with people and intelligent agents in both synchronous and asynchronous ways that is supportive of multi-tasking.

Multi-tasking

Different communication activities have different efficiencies. The following table shows rough estimates of average speed of different communication activities:

Communication Activity	Speed (words/min)	Average time per person per 100 words (mins)	Relative (%)	Relative (ratio)
Talking/Listening	190	0.52	57%	4
Reading	333	0.30	100%	7
Typing (keyboard)	50	2.00	15%	1
Typing (SMS)	10	10.00	3%	0.2

Social trends in the workplace include rising productivity and rising salaries. This points to an increasing need for employees to be more productive by multi-tasking: carrying out more than one task at a time. All other things being equal, a user of a communication system is likely to prefer a communication system that is designed to support multi-tasking.

Some tasks can be carried out at the same time whilst others cannot. The following table shows a set of 'rules of thumb' for which communication activities can be multi-tasked. Each mode shows a pair of tasks that can be performed together. It is assumed that three tasks cannot be performed simultaneously by most people.

Mode	Talk	Listen	Read	Visual Media (Silent)	Type/ write	Manual eg making tea	Sign / face / body language
1	x					x	
2	x						x
3		x	x				
4		x		x			
5		x			x		
6		x				x	
7		x					x

Modes 3 and 5 in the table above are perhaps the most common modes of multi-tasking on a personal computer between different task types. An integrated avatar user interface system should support reading and typing tasks whilst listening.

Another type of multi-tasking is time efficiency of verbal communication. Whilst in an avatar user interface session, the participant should be able to carry out other voice tasks in the periods when the conversation of the session is not important to him. The following voice tasks are possible whilst in an avatar user interface session:

- listen to voice-mail
- speak voice-mail
- make a voice call
- receive a voice call
- interact with conversational intelligent avatar agents
- interact with user interfaces

Voice task functions

Some functional considerations are important for voice tasks:

- mixing of conference audio with the incoming voice signal so that participants can be passively aware of what is happening in the conference. An example is someone asking you a question whilst you are on another task, in which case you are likely to hear "What do you think YourName?" and react appropriately

- easy to use switchboard between voice tasks. Multi-tasking requires speed and efficiency in switching between synchronous voice tasks such as putting a party on hold
- rapid directory look-up of a person, indication of whether a person is logged on / active on his personal computer and automatic dialling of a voice call.
- visual status of voice tasks that are active or on hold
- mixing of voice functionality and text functionality
- ability to switch off direct voice access to yourself giving you time to think or just have a break; voice calls can be diverted to voice mail for listening to them later

Always-on

Nowadays, most business internet connections are always-on rather than dial-up and this trend is likely to continue. Immediate Messaging (IM) is often present on the desktop in businesses all the time; with IM people can respond to incoming messages immediately. People spend ever more of each day at their personal computers. Many employees listen to music through headphones while they work.

There are times when people do not wish to be interrupted by IM or by unplanned voice calls. In these times, IM can be switched to e-mail and voice calls to voice-mail.

Communication types

Communications may use any type of media or any combination of multiple media. The term multi-media is used to cover all types of media such as but not limited to text, voice, video, image, animation and avatar.

The following communication types are usually available in an Avatar User Interface System by way of example but an Avatar User Interface System is not limited to these communication types. Synchronous communication is when a communication is usually received in real-time and often responded to in real-time. Asynchronous communication is when a communication is usually received after a delay.

Synchronous

- avatar user interface session / avatar conference
- voice call / avatar call
- Immediate Messaging
- video
- Whispering

Asynchronous

- voice-mail / avatar voice-mail / video-mail
- e-mail

An avatar call is when the avatar 5 of the user 17 appears in the Meeting Room Media window 50 whilst the synchronous voice call takes place. An Avatar voice-mail is when the avatar 5 of the user 17 appears in the Meeting Room Media window 50 whilst the asynchronous avatar voice-mail is being played back.

Avatar User Interface

User interfaces have evolved over the course of computer history. They have progressed from a bank of switches and lights through to sophisticated windows interfaces for information interaction and communication tasks. With the very recent advent of photo-realistic avatars and the maturing of voice processing technology, a new form of user interface is possible.

It is a purpose of this avatar user interface system invention to disclose a new form of user interface for interacting with people, information, entertainment and avatar agents.

Figure 42 is a representation of an example of the displayed avatar user interface 260 in this fifth embodiment. The new switchboard avatar user interface functionality 268 is added to the avatar session user interface 10 shown in Figure 12 such that both sets of functions are integrated and easily accessible through the same user interface hardware 3.

The switchboard avatar user interface functionality includes:

- a buddy list 240 with data for each buddy such as name 244 and facial icon 243
- buddy list buttons such as add buddy 247, edit buddy 248 and delete buddy 249
- a switchboard 241 with numbered events 252 including live sessions with data for each session party such as name 244 and facial icon 243, live conferences with conference name 253 and with data for each conference attendee such as name 244 and facial icon 243, streaming media channels 254 such as music, radio and TV and voice mails 255
- a status bar 250 with a message 251 and session control buttons such as start session to new party 242, end session 246 and whisper to a party on a session 245

Multiple session servers

In a working day, millions of meetings and calls take place. It is technically challenging that all avatar user interface sessions 252

are served by a single session server 1. Additionally, it is likely that several companies will compete in the avatar user interface system marketplace, each company having one or more session servers 1.

Figure 43 is a block diagram of a multi-session server system. It shows a personal computer 3 with an avatar user interface 260 connected via a network 2 to two or more session servers 1. Different live sessions 252 may take place on different session servers 1. Protocol converters 301 are resident at different places on the system.

Large economic benefits provide strong commercial forces for players in a market to agree standards. It is likely that a standard avatar interface protocol 300 between session servers 1 and avatar user interfaces 260 will be agreed for avatar user interface systems 261. Eventually this could form a global standard.

It is a purpose of this avatar user interface system invention that an avatar user interface 260 on a personal computer 3 can simultaneously be connected to multiple sessions on a plurality of session servers 1 using a standard avatar interface protocol 300.

Multiple protocols

If two or more standard avatar interface protocols 300 are used, protocol converters 301 can convert between the protocols in real time. The protocol converters 301 can be situated on the network 2 or within the session servers 1 or within the avatar user interface 260 or any other suitable place.

It is a further purpose of this invention that where two or more standard avatar interface protocols 300 are used, that protocol converters 301 can convert between the protocols in real time.

SIXTH EMBODIMENT

It is a purpose of this sixth embodiment that the invention is not limited to the displayed avatar user interface 260 running in a browser window 21, but that it can run as a stand alone application.

Figure 44 is a block diagram of the displayed avatar user interface 260 on the display device 264 being driven by the avatar user interface software application 262 running stand-alone in memory 345 on the personal computer 3.

SEVENTH EMBODIMENT

It is a purpose of this seventh embodiment that the displayed avatar user interface 260 might be for a digital exhibition in which there are many virtual stands representing different organisations on which information about their products and services is accessible.

Currently, there is no effective means of visiting an exhibition virtually. There is also no virtual means for simultaneously talking to a salesman, viewing his photo-realistic avatar and seeing information on the company's products and services.

Figure 45 is a representation of an example of the displayed avatar user interface 260 containing the switchboard avatar user interface functionality 268, the avatar session user interface 10 and the exhibition user interface functionality 280. The exhibition user interface functionality 280 includes an exhibitor list 281 with different organisations 282 that can be selected. Pressing a browse button 283 enables the user 17 to enter a 3D meeting room media window 50 of the selected organisation in which information media about the organisation's products and services is available for browsing by the user 17. Pressing a contact button 284 enables the user 17 to call a representative of that organisation into the organisation's 3D meeting room media window 50. The representative can be an actual person with his own avatar or an intelligent agent avatar. As on a physical exhibition stand, multiple users 17 can be present with one or more representatives of the organisation in the same exhibition 3D meeting room media window 50.

Whilst browsing in the 3D meeting room media window 50, a user 17 may:

- see objects representing products 286
- navigate by pressing on the object 286 or pressing buttons on the navigation bar 287

- view it by pressing a button 288. The user's avatar can pick the product 286 up and turn it around if the product is of a suitable size. Alternatively, the product can be rotated by the user.
- buy it by pressing a button 285
- be taken on a tour around the company's products 286 by an intelligent agent avatar 5 if the user 17 presses the button 289.

This embodiment is not limited to the functions disclosed here but covers any function from an actual exhibition that can be implemented virtually.

There are many advantages of virtual avatar exhibitions as disclosed in this embodiment including:

- interactivity with the salesman using voice to communicate
- ability to browse a 3D virtual stand without being approached by a salesman
- not having to spend time and incur cost travelling to actual physical exhibition locations
- not missing an exhibition due to a schedule clash

It is a purpose of this seventh embodiment to disclose a process wherein users communicate in virtual exhibition means comprising the following steps:

- a user navigates in a virtual exhibition stand of a company;
- the user views and interacts with virtual objects representing products;
- optionally the user communicates remotely with a real sales representative;
- optionally the user communicates with an intelligent agent avatar;
- optionally the user views presentations;
- optionally the user buys the product.

EIGHTH EMBODIMENT

It is a purpose of this eighth embodiment that an avatar agent 5 is driven by an intelligent agent and not by the user 17.

Avatar agents

Avatar agents are photo-realistic avatars driven by intelligent software agents rather than people. An avatar user interface system that will provide the benefits of avatar agents to people does not exist.

Intelligent agent

Figure 46 is a block diagram of an avatar agent hosting system and intelligent agent software in accordance with an eighth embodiment of the present invention. It shows intelligent agent software unit 320 on an avatar agent hosting server (AAHS) 321 running with AAHS management software 322 stored in memory 348 driving an avatar agent 5 in an avatar user interface window 260 on the display device 264 of a personal computer 3. The AAHS management software 322 manages one or more intelligent agent software unit 320 running concurrently on the AAHS 321. Alternatively, in a second client-server system architecture, the intelligent agent software unit 320 may be running in memory 344 on the avatar hosting server (AHS) 4. Alternatively, in a peer to peer system architecture, the intelligent agent software unit 320 may be running in memory 345 on a personal computer 3.

The identity 275 of the avatar agent 5 is usually the same as for the intelligent agent software unit 320. The identities of the avatar agent 5 and the intelligent agent software unit 320 could also be different, in which case the avatar agent identity number would have to indicate on which avatar agent hosting service the avatar agent is resident.

The intelligent agent software unit 320 can perform synchronously or asynchronously. It can communicate by outputting marked-up text 327 or audio voice 185. It contains artificial intelligence software 323 and a database of knowledge 324. It may also have access to further databases of knowledge 324 via the network 2. For voice communication it includes a speech recognition engine 182 and an agent text to speech engine 326.

The intelligent agent software unit 320 can generate events 81 that are incorporated as mark-ups in the marked-up text 327 or output from

the agent text to speech system 326. The events 81 that go to the software director 80 cover such aspects as emotions and gestures.

The actions 83 of an avatar agent 5 usually exhibit better realism than the actions 83 of an avatar 5 driven from voice 185 because the intelligent agent software unit 320 has more knowledge for generating events 81 than can be extracted from analysis of the live voice stream 185 of a user 17.

The intelligent agent software unit 320 can represent itself visually with an avatar 5 that does not have the identity of a real person. Each avatar agent 5 is driven by one intelligent agent software unit 320. The avatar agent 5 of the intelligent agent software unit 320 may be a parameter avatar 232, or it may be edited to look like a photo-realistic avatar 238 of a real person or it may be based on images taken of a real person with whom that person's identity is not associated.

The intelligent agent software unit 320 speaks through an agent text to speech engine 326 using impersonation parameters 325 that makes the voice 185 emit a characteristic profile. An example of a characteristic voice profile 184 is a middle-aged Scottish woman. The avatar agent 5 is impersonating a middle-aged Scottish woman. The impersonation parameters 325 are of two types: voice impersonation parameters 331 and action impersonation parameters 332.

Agent impersonation of a person

The agent avatar 5 can represent a real person 17 and use the photo-realistic avatar 238 of that real person 17. The impersonation parameters 325 can be the personalised voice profile of that particular person 17. In this way the avatar agent 5 can represent the real person 17 by looking like that person and sounding like that person whilst that real person is unavailable.

Generating impersonation parameters

Figure 47 is a block diagram of an apparatus for generating impersonation parameters. To obtain a high quality of recording, a person's 17 voice and movements may be recorded in a room 330

insulated from a noisy environmental location 273. Video 336 is recorded from at least one camera 29 and audio 185 is recorded using a microphone 12 of a person 17 reading known text 189 on a screen 264 of a personal computer 3. The impersonation parameter generation software 331 running in memory 345 on the personal computer 3 processes the video 336 and audio 185 to generate a set of impersonation parameters 325. The impersonation parameters 325 are of two types: voice impersonation parameters 331 and action impersonation parameters 332. The voice impersonation parameters 331 are generated by processing the audio of the known text 189. The action impersonation parameters 332 are generated by processing the facial movements as the words in the known text 189 are spoken and as emotions are used.

Using impersonation parameters

The intelligent agent software unit 320 generates marked up text 327. The marked-up text 327 is processed by the agent text to speech engine 326 using the voice impersonation parameters 331 of the person 17 to modify an existing speech database 328 by speech synthesis. The voice 185 emitted by the agent text to speech engine 326 sounds like that of the person 17. The marked up events in the marked-up text 327 are modified by the action impersonation parameters 332 to produce gesture action events 81 that use characteristic gestures that the person 17 normally uses when speaking.

The simplest application of avatar agent impersonation of a real person would be in a personalised answer-phone application. The intelligent agent software unit 320 may know who is calling and their access level to the real person's information. It may also know what activity the real person is currently involved in and when it is due to finish. It answers the avatar call with an appropriately personalised message. For example: "Hi John, I'm in an avatar session until 11.00, please leave an avatar-mail." The caller, John, will recognise the voice and see the avatar as if it were the real person he was calling. In applications requiring more intelligence from the intelligent agent software unit 320 such as a personal assistant application, more advanced bi-directional communications can take

place in which, for example, arrangements can be 'pencilled in' involving diaries.

Avatar agents in this embodiment of the avatar user interface system invention will provide benefits to people in a wide range of applications, including as:

- call centre personnel: users can interact with an organisation via virtual agents instead of expensive call centre personnel; fields include: account payments, technical support, changing service levels
- sales representatives: users can discuss potential purchases with virtual agent sales representatives
- real estate: home purchasers can be shown round virtual 3D replicas of homes on the market by a virtual real estate agent
- entertainers: avatar agents will become performers in shows customised to the user's desires; what was a child's television programme becomes an interactive, personalised entertainment led by an avatar agent
- advisers: people can consult an avatar agent specialist for advice; fields include: independent financial advice, style of clothing, selection of make-up, dieting, fitness, sports, psychology, psychiatry, cooking
- newscasters: virtual newscasters will be able to read the news that you want when you want
- housekeepers: virtual agents will provide: management of the home network, automatic call out of home service personnel such as heating system technicians, automatic reordering of home consumables such as lavatory rolls, entry to trusted persons, interfacing with home residents via an avatar user interface system
- personal assistants at work: a virtual agent will become your personal assistant managing tasks including booking meetings, taking messages, making travel arrangements, carrying out research
- teachers: virtual tutors will time and pace e-learning courses to suit you and maximise your rate of skill acquisition; fields include education at all levels, music in which the avatar agent music teacher records and analyses a student's playing of an instrument, languages in which the avatar agent language teacher

can correct pronunciation and lead a discussion in the foreign language

- representatives of yourself: an avatar agent can represent you when you are off-line, participating in interactions with people and other avatar agents; looking like you and sounding like you

Eventually, people will interact with avatar agents as they do with other people. It will be very difficult to distinguish between avatar agents and avatars driven by people. Where such communication is remote (not face to face), it is likely that the same interface will be used for conversing with people as agents.

It is a purpose of this eighth embodiment to disclose a process wherein generic action impersonation parameters are defined for a communication context comprising the following steps:

- recording a corpus of videos of the communication context;
- processing the corpus by a trained person along a timeline to produce an annotated timeline with actions of each communication context participant related to a number of parameters;
- analysing the annotated timeline by a trained person to produce a type definition of each action impersonation parameter and a set of rules that can be incorporated into a finite state machine for the communication context.

It is a further purpose of this eighth embodiment to disclose a process wherein personal action impersonation parameters for a particular person are generated using an action impersonation generator/editor means involving manual input by a user comprising the following steps:

- in the first step, the user makes selections from a number of sets of generic action impersonation parameters at a high level;
- in the second step, the user edits the selections at a lower level;

wherein the second step is optional and the user may or may not be the person for whom the personal action impersonation parameters are generated.

It is a further purpose of this eighth embodiment to disclose a process wherein personal action impersonation parameters for a particular person are generated automatically using an action impersonation generator/editor means comprising the following steps:

- in the first step, video recordings are made of the person carrying out a number of defined actions;
- in the second step, the action impersonation generator/editor automatically analyses the video recordings to generate a set of personal action impersonation parameters.

It is a further purpose of this eighth embodiment to disclose a process wherein a software director uses voice impersonation parameters defined for an avatar to generate speech from text using text to speech engine means for an avatar such that the avatar speaks recognisably like the person it represents comprising the following steps:

- intelligent agent software unit means generates the text;
- text to speech engine means converts the text to speech;
- the speech is played on the computing appliance.

It is a further purpose of this eighth embodiment to disclose a process wherein voice impersonation parameters are defined for an avatar of a particular person comprising the following steps:

- recording the person speaking predefined text;
- processing the recording using impersonation parameter generation software;
- the impersonation parameter generation software outputting the voice impersonation parameters for that person;
- storing the voice impersonation parameters in the avatar.

It is a further purpose of this eighth embodiment to disclose a process wherein after a voice communication by a user, a speech recognition engine means processes the voice communication comprising the following steps:

- a user generates a voice communication by speaking;
- a speech recognition means processes the voice communication and outputs text;

- the text is sent to any intelligent agent software units involved in the session.

It is a further purpose of this eighth embodiment to disclose a process wherein a user speaks in a first language and an intelligent agent software unit operates in a second language such that text is translated by translation engine means comprising the following steps:

- a user generates a voice communication by speaking in a first language;
- a speech recognition means that operates in the first language processes the voice communication in the first language and outputs text in the first language;
- the text in the first language is translated by translation engine means into text in a second language;
- text in the second language is sent to any intelligent agent software units involved in the session capable of processing text in the second language.

It is a further purpose of this eighth embodiment to disclose a process wherein a user understands a first language and an intelligent agent software unit operates in a second language such that text is translated by translation engine means comprising the following steps:

- an intelligent agent software units generates text in a first language;
- the text in the first language is translated by translation engine means into text in a second language;
- text to speech engine means converts the text in the second language to speech in the second language;
- the speech in the second language is played to the user using loudspeaker means.

NINTH EMBODIMENT

It is a purpose of this ninth embodiment that the avatar user interface system 261 may be used for biometric security applications at locations such as airports or military installations.

There is an increasing need for security systems based on biometric identification at airports to combat terrorism and in many other security applications. Currently, a biometric security system based on photo-realistic avatars of people does not exist.

Biometric security

Figure 48 is a block diagram of the avatar user interface system with extended security functionality in accordance with a ninth embodiment of the present invention. It shows a person 313 passing a security checkpoint 314. The person's identity 275 is contained on an identity source 310 such as a smart card carried by the person 313 or an implant in the person 313. The person's identity 275 is read from the identity source 310 by an identity source reader 311 attached to a personal computer 3. Identity processing software 312 in memory 345 on the personal computer 3 calls up the avatar 5 corresponding to the identity 275 from the avatar hosting service 4 over the network 2. The avatar 5 corresponding to the identity 275 is displayed in the avatar user interface window 260 on the display device 264 of the personal computer 3. A security user 17 who is usually a security guard, can visually compare the person 313 to the avatar 5 corresponding to the identity 275 presented by the person 313. If the person 313 and the avatar 5 are not similar then the security user 17 can stop the person 313 for questioning. To enhance security, the network 2 and avatar hosting service 4 may be private to the organisation conducting the security check.

In a more automated method of conducting security checks, a camera 29 attached to the personal computer 3 takes images 19 of the person 313. Image processing and comparison software 315 in memory 345 on the personal computer 3 can automatically compare the images 19 to the avatar 5 corresponding to the identity 275 presented by the person 313. If the image processing and comparison software 315 finds a significant discrepancy between the images 19 and the avatar 5 corresponding to the identity 275 presented by the person 313, then the security user 17 is alerted. Image processing and comparison software 315 are well known to those skilled in the art; however, increasing the accuracy of such software is still a research area.

Where a discrepancy is found, the images 19 may be sent to remote image processing and comparison software 315 in memory 344 on the avatar hosting server 4 which will compare the images 19 with a database 318 of images of known people because they are a security risk or employees or for any other reason. If the remote image processing and comparison software 315 on the avatar hosting server 4 makes one or more possible matches then these possible matches are communicated to the security user 17 via the avatar user interface window 260 on the display device 264. The remote image processing and comparison software 315 and database 318 are not limited to being resident on the avatar hosting server 4 but may be resident on any server accessible via at least the network 2.

An intelligent agent software unit 320 resident in memory 345 on the personal computer 3 represented by an avatar 5 in the avatar user interface window 260 on the display device 264 communicates with the security user 17. The intelligent agent software unit 320 may generate communications to the security user 17 relating to the advisable actions to be taken depending on the results of any comparisons made by the image processing and comparison software 315. The intelligent agent software unit 320 may respond to communications from the security user 17 such as requests for further comparisons.

Certainty in verifying identity can be increased by combining the results of two or more biometric devices. In addition to the camera 29 comparing to the avatar 5, a biometric device 316 connected to the personal computer 3 could measure another part of the person and compare it to reference biometric data 317 linked with the avatar 5 via the identity 275. Typical biometric devices include fingerprint scanning, iris scanning, hand scanning, face recognition and voice pattern recognition.

It is a purpose of this ninth embodiment to disclose a security process comprising the following steps:

- a person providing an identity source that is read by an identity source reader;
- retrieving the avatar whose identity matches the identity in the identity source;

- displaying the avatar;
- a security user visually comparing the avatar with the person.

It is a further purpose of this ninth embodiment to disclose a largely automated security process comprising the following steps:

- the person providing an identity source that is read by an identity source reader;
- retrieving the avatar whose identity matches the identity in the identity source;
- extracting the avatar biometric data from the avatar;
- a biometric device scanning part of the person to provide scanned biometric data;
- comparing the scanned biometric data with the avatar biometric data;
- if the scanned biometric data does not match the avatar biometric data then alerting the security user;
- displaying the avatar to the alerted security user;
- the alerted security user visually comparing the avatar with the person.

This embodiment is not limited to the disclosure provided. For example, the intelligent agent software unit 320 might be resident on an avatar agent hosting server (AAHS) on the network 2 instead of on the personal computer 3. Two or more security checkpoints 314 at one location may be connected to a single personal computer 3. In a large establishment, multiple security checkpoints at multiple locations may be wired to multiple personal computers 3 in one or more security rooms monitored by multiple security users 17.

This embodiment of the avatar user interface system invention has significant utility. It can support a security guard in making a quick visual verification that the person showing an identity is actually the person to whom the identity belongs. In a more automated form, it can alert a security guard when a discrepancy between the person going through a security checkpoint and his avatar is detected.

TENTH EMBODIMENT

It is a purpose of this tenth embodiment that the avatar user interface system 261 is used for interactive computer games.

On-line interactive computer games do not exist where the user is represented by a photo-realistic avatar 238 of himself. On-line interactive computer games do not exist where avatars 5 can exhibit lip synchronised animation in real-time to voice transmitted over a network 2.

Games

Figure 49 is a block diagram of an avatar user interface system 261 for interactive computer gaming in accordance with a tenth embodiment of the present invention. Users 17 interact with their personal computers 3. A session server 1 handles the voice mixing between users 17. An avatar hosting server 4 hosts the avatars 5 of the users 17 which are sent to the personal computers 3. A Game Hosting Server 370 hosts the game software 371, the state 372 of the game and billing software 237 in memory 374. A network 2 connects the servers and personal computers. If the game involves avatar agents, one or more avatar agent hosting servers 321 may serve the avatar agents 5.

Special game interface equipment 373 may be attached to the personal computer 3 containing sensors to detect the movements of the user and feedback devices to stimulate the user's senses.

The computer games industry is clearly structured with a number of game genres such as roll playing games (RPG), sports including football and wrestling, car racing, God simulations, strategy games, board games and first-person fighting.

Some of these genres have found a place in on-line gaming in which users play the game with each other over a network. The game is hosted on a game server that is also on the network.

It is a purpose of this avatar user interface system invention that a new genre of communicative avatar on-line game may be built using the avatar user interface system 261 that was not possible before. Users

17 see each other in the virtual environment of the game software 371 as photo-realistic avatars 5. During the game, users 17 can communicate with each other by voice as if they were in the same room.

In some game designs, the user 17 may navigate his avatar 5 through the virtual environment of the game using normal personal computer input devices such as mouse and keyboard. Examples of this are environments where users navigate towards other people's avatars to meet them.

In other game designs, the actions of the user's avatar are generated by the software director 80 in reaction to events. As an example, in a game of chess, the user 17 may move one of his chess pieces from one square to another and the software director 80 will show the user's avatar 5 picking up a piece and moving it from one square to another.

In an on-line game 371 with a shared virtual environment, the state 372 of the game must be maintained on the game hosting server 370. In this way, the shared virtual environment of the game 371 is the same for all users 17 at all times because there is only one state. The only time that there are differences is if there are delays or lags on the network 2. However, state discrepancies caused by the software director 80 playing actions in anticipation of what will happen in the game 371 whilst waiting for the new state 372 to be synchronised over the network 2 delays can be quickly corrected by the software director 80 when the new state 372 arrives at the personal computer 3.

It is readily apparent that there are advantages of this embodiment when during an online interactive computer game either photo-realistic avatars 238 can be viewed or avatars 5 can be lip synchronised or both. The advantages include combined audio and visual recognition and suspension of disbelief such that the user finds the game compelling and the sessions are longer.

It is a purpose of this tenth embodiment to disclose a process wherein users communicate in an interactive game hosted on a game hosting server comprising the following steps:

- a first user interacts with the game, navigates around the 3D game scene and views the avatar of a second user;
- the second user interacts with the game, navigates around the 3D game scene and views the avatar of the first user;
- the first user communicates by speaking;
- the second user hears the first user and views the avatar of the first user in lip synchronisation with the first user's speech;
- the second user communicates by speaking;
- the first user hears the second user and views the avatar of the second user in lip synchronisation with the second user's speech.

ELEVENTH EMBODIMENT

It is a purpose of this eleventh embodiment that the avatar user interface system 261 is used in immersive virtual reality (VR) environments.

There is a variety of immersive VR systems. These include but are not limited to VR headsets and caves. A person can wear a VR headset in which his view of the physical environmental location 273 is replaced by viewing a display apparatus on which the virtual environment is displayed. A person can enter into a cave, that can be generally defined as a partially or fully enclosed physical space in which the display area is large. The person sees a virtual environment that has been projected onto the walls, floor and ceiling of the room.

An immersive VR system does not exist in which people are represented by photo-realistic avatars of themselves. Nor does an immersive VR system exist where people's movements can be motion tracked and used to drive photo-realistic avatars of themselves in other locations.

Immersive VR

Figure 50 is a schematic of an avatar user interface system 261 for a six-sided cave 350 in accordance with an eighth embodiment of the present invention. The six faces of the cave 350 are illuminated by six back projectors 352. The six back projectors 352 are connected by one or more cables 354 to a computer 355. The computer 355 contains an avatar player engine 210 and a cave display system 357 in memory

345 and is connected to a network 2. The avatar player engine 210 generates a 3D virtual environment 356 containing avatars 5 that usually changes over time. The avatar player engine 210 transmits the 3D virtual environment 356 to the cave display system 357. The cave display system 357 generates the digital projector images 353 and transmits them to the back projectors 352. The six faces of the cave 350 are fabricated from a material that permits back projection such that the six projected images 353 are visible to the user 17 from inside the cave 350. Each projected image 353 is a sequential stereo pair from which a 3D effect can be experienced. A user 17 wearing shutter glasses 351 is inside the cave 350. The shutter glasses 351 combine the stereo pair image 353 displayed onto each wall to form a 3D virtual environment 356. The 3D virtual environment 356 appears to stretch from right next to the user 17 to many hundreds of metres away. The experience is vivid and a strong sense of presence in the virtual world is experienced by the user 17. The user 17 sees a 3D virtual environment 356 with an avatar 5. The user 17 can see an avatar 5 in 3D when the user 17 is facing in the direction of the avatar 5. If the avatar 5 is central in the cave 350, the user 17 can walk through or around the avatar 5, turn and see the avatar 5 from a different viewpoint. The avatar 5 can move in the virtual environment 356 relative to the user 17.

It is often possible for several people to be in a cave simultaneously. In a cave 350, the user 17 can see parts of himself 17 such as his legs 358 and arms 359, the people with him in the cave, the virtual environment 356 and the avatars 5.

This eleventh embodiment is not limited to a cave with six sides, a physical space with a display of one or more sides can be used. Conventional displays such as a monitor or plasma screen can be used. Shutter glasses are one method of converting images into a 3D environment, but this invention can incorporate a wide range of 3D display technologies.

Networked VR

One or more users 17 in a cave 350 at a first location can be connected via a network 2 to one or more other users 17 in another

cave 350 at a second location such that all the users 17 appear as avatars 5 immersed in the same 3D virtual environment 356. It is advantageous for users at one cave location to see the movements of the users at the other cave location. For this, a motion capture system is required at each location.

Figure 51 is a schematic of an avatar user interface system 261 for two caves 350 connected by a network 2. A motion capture system 368 is integrated into the cave 350 at location 1. The motion capture system 368 comprises four cameras 362 viewing the internal area of the cave 350 connected by a cable network 365 to a computer 363 running motion capture software 364 in memory 345. The user 17 wears a suit 360 to which infra-red emitters 361 are attached. As the user 17 moves around in the cave 350, the motion capture software 364 on the computer 363 calculates the motion 369 of the user 17. The motion 369 is sent to the cave 350 at location 2 and the motion 369 is played on the photo-realistic avatar 5 of the user 17.

There are many types of motion capture system and this invention is not limited to the type disclosed. For example, the motion capture system might be passive and not require the user to wear a seat with active emitters.

As an alternative to a cave, a user 367 may wear a VR headset 366 whilst moving inside the motion capture system 368.

As an alternative to cabling 354 and or 365, wireless networks could be used.

Some of the avatars 5 might be avatar agents 5 driven by intelligent agent software unit 320 rather than users 17. In this way, agents and users can mingle and interact in a 3D virtual environment 356 without it being immediately obvious which avatar 5 is driven by an agent or a user.

This invention is not limited to participants in just two locations being immersed in the same 3D virtual environment 356. Three or more

locations may be used. At each location, there could be a cave or the user could use a VR headset.

At some locations there may not be motion capture so that the user could not be seen to move around. Where the user is viewing only he could either be invisible to the other users or visualised in a fixed position either standing or sitting. In this instance, the user could use an avatar user interface as disclosed in the First Embodiment.

During networked immersion, each user 17 can wear a headset 11 for audio communication with the other participants.

There are significant advantages of this embodiment of the avatar user interface system. Fundamentally, this embodiment discloses means by which the most realistic immersive VR experience can be achieved and will thereby achieve a high sense of presence in the session. Applications of this eleventh embodiment, cover entertainment, communication and collaborative work; but this embodiment is not limited to these applications.

It is a purpose of this eleventh embodiment to disclose a process wherein users are present in Cave means with motion capture systems means comprising the following steps:

- the motion capture system means records movements of a first user in a first Cave means;
- the recorded movements are sent with acceptable lag from the first Cave means to a second Cave means;
- an avatar of the first user is displayed in the second Cave means such that the movements of the avatar duplicate the movements of the user in space;
- a second user wearing shutter glasses or similar immersive 3D viewing means in the second Cave means views the movements of the avatar of the first user as if the first user were physically in the second Cave with the second user.

TWELFTH EMBODIMENT

It is a purpose of this twelfth embodiment that the avatar user interface system 261 may be connected to exercise equipment.

Health and training

There is an increasing consumer demand for healthier lifestyles. People buy exercise equipment such as rowing and running machines for use at home. However, many people lack the motivation to use them on their own.

Figure 52 is a schematic of an avatar user interface system 261 comprising two exercise stations 414 connected together by a network 2. An exercise station 414 comprises a piece of exercise equipment 410 used by a user 17 wearing a pulse rate gauge 415 with a processor 411 connected by a cable 413 to a personal computer 3 running exercise equipment interface software 412 and avatar user interface software 262 in memory 345 with a display device 264 showing an avatar user interface 260 viewed by the user 17.

Many items of exercise equipment come with a built-in processor and a connection to a personal computer such that the personal computer can monitor and or control the exercise equipment. Examples of parameters monitored from the exercise equipment might be speed, strength setting, energy dissipation rate, user pulse rate and cumulative energy dissipated. An example of a parameter that might be controlled is the strength setting of the exercise equipment.

The two users can share their exercise as a social experience. A first user 17 can see the avatar 5 of the second user 17 in his avatar user interface 260 in a scene showing the avatar 5 of the second user using a virtual exercise equipment 410. As the exercise equipment 410 of the second user 17 is moved by that user, the exercise equipment interface software 412 monitors the movements of the exercise equipment 410 and sends them over the network 2 to the avatar user interface software 262 on the personal computer of the first user 17. In this way, the first user sees the avatar 5 of the second user moving on the virtual exercise equipment in the avatar user interface 260 in substantially real time compared to the actual movements of the second user. If the second user stops using the exercise equipment, then the first user will see almost immediately that the avatar of the second user has stopped using the exercise equipment. During the

avatar user interface system interaction, the two users may talk to each other using the headsets 11.

The scope of this twelfth embodiment is not limited to the disclosure above. As will be understood by persons skilled in the art, the connection between the exercise machine 410 and the personal computer could be wireless rather than a cable 413. The display device 264 and the personal computer 3 might be built into the exercise machine 410 which connects to the network 2. The headset 11 may be connected to the personal computer 3 by wireless rather than a cable. Loudspeakers may be used instead of headphones. Other biometric devices may be worn by the user 17 in addition to the pulse rate monitor. The exercise equipment interface software 412 may correlate performance of each user over a number of sessions and generate statistical data to track increases in fitness. The wearing of a pulse rate gauge 415 is optional.

This twelfth embodiment is not limited to two users but three or more users may be connected simultaneously. One user 17 may be a personal trainer for another user 17 and use the avatar user interface system to both monitor and encourage the first user. A personal trainer could train several users simultaneously. Users may compete against each other on certain parameters such as speed, strength and endurance. International virtual competitions may be held with their appearance in the avatar user interface system being similar to that of a televised sports event. A user 17 may be a medical doctor who can monitor remotely the health of a user 17 who is a patient. An avatar intelligent agent software unit 320 may take the role of a user or personal trainer or doctor or any other professional such as a sports therapist.

Furthermore, this twelfth embodiment may be combined with features from the fourth embodiment such that two or more people at one location can exercise together whilst being in contact with one or more other people at one or more other locations.

This twelfth embodiment of the avatar user interface system invention enables a person who is in one location to carry out a physical

activity whilst in virtual contact with one or more people in other locations. Advantages of this embodiment include: time and cost saved travelling by each user to an agreed location where they can exercise together, increased motivation by exercising whilst in virtual contact and not needing to dress up to be seen in public.

It is a purpose of this twelfth embodiment to disclose a process wherein users communicate whilst exercising on exercise station means comprising the following steps:

- a first user using a first exercise station means;
- a second user using a second exercise station means;
- the first user viewing the avatar of the second user using a virtual exercise station;
- the second user viewing the avatar of the first user using a virtual exercise station;
- the first and second users communicating by voice;
- optionally the first and second users viewing performance data generated by the first and second exercise station means;

optionally any user being able to see if the other user has stopped exercising.

THIRTEENTH EMBODIMENT

It is a purpose of this thirteenth embodiment that the avatar user interface system 261 may be used for practicing and planning.

Currently, no avatar user interface system means exist for a person to practice or plan something virtually, either with other people or with avatar agents or both.

Practicing might cover exercises for learning a new skill, preparing for delivery of an event or planning an event. Examples of applications that require practicing include: language learning, learning touch typing, delivering a presentation, public speaking, playing music, rehearsing a play, overcoming a fear such as that of public speaking by practicing in a virtual environment, planning the choreography of a ballet and planning the direction of an event.

Practicing using the avatar user interface system 261 involves the person practicing generating output into the avatar user interface system 261 by means of voice, camera, keyboard, mouse or other specialised peripheral. This input may be fed to another person or an agent where it is processed and feedback is given to the person practicing. Feedback may be verbal or visual. Emote keys may be used by the person feeding back such that a person's avatar can visually show pleasure, displeasure, comprehension, confusion and other emotions.

In the case of planning, a person planning will create a plan. This can be done collaboratively with others in synchronous or asynchronous ways. Synchronous planning will involve real-time interactions between users. Asynchronous planning might involve one person creating a plan such as a choreography for a ballet and others feeding back at a later time. In this case, a set of tools and props will usually be required for the application being planned.

FOURTEENTH EMBODIMENT

It is a purpose of this fourteenth embodiment that the avatar user interface system 261 has an Avatar Virtual Environment (AVE) as the background to the display device 264 and that the desktop 423 is present and usable on a virtual computing appliance 421 within the AVE.

Figure 53 is a schematic of the display 264 of an avatar user interface system 261 with an avatar virtual environment (AVE) 420 as the background in accordance with this fourteenth embodiment. A virtual computing appliance 421 with a virtual computing appliance display 422 is present in the AVE 420. The desktop 423 of the PC 3 is shown on the virtual computing appliance display 422. The virtual computing appliance 421 is not always visible in the AVE 420 because visibility depends on whether it falls within the field of view of the virtual camera being used at that instant.

Avatar Virtual Environment (AVE)

Existing PC operating system user interfaces 20 are largely based upon the Windows concept such as the Microsoft Windows XP operating system.

This user interface concept is referred to as the windows user interface. The windows user interface usually occupies the whole of the display area of the display device 264. The windows user interface usually consists of a desktop 423 background covering the whole display area and may have one or more windows open on top of the desktop 423. Any one window may be opened fully to cover the whole desktop 423.

This avatar user interface system invention includes one or more windows containing an Avatar Virtual Environment (AVE) 420. An AVE is a photo-realistic virtual environment with photo-realistic avatars in it. The avatar user interface system of the First Embodiment uses avatar conference windows 23, 24 and 25, which are AVE windows, open in the context of the windows user interface. Controls such as control buttons 27 are situated outside the avatar conference window.

Frames, presence and cognitive jolt

A person often has several frames of interaction around him. These might include other people in the room, the computer's desktop display, various active applications on the display, music and a telephone. At any one time, the person's bandwidth of consciousness is spread between one or more frames. Although the term presence does not have a generally accepted definition, applications with high presence are applications in which the person tends to be very immersed such that awareness of other frames is only peripheral.

Normally, there is a strong cognitive jolt in switching between frames. Someone may call you by name when you are playing a computer game and it may break your concentration with a jolt. The design of the windows user interface minimises cognitive jolt when moving between application windows. Currently, an AVE 420 operating as a window on a desktop 423 tends to result in a low feeling of presence because there is no meaningful metaphor between the AVE window and the neighbouring applications. There is also a high cognitive jolt when transferring between the AVE window and a neighbouring non-AVE desktop window.

By placing the desktop 423 on a virtual computing appliance display 422 within the AVE 420, there is a continuous metaphor and lower cognitive jolt as the user 17 transfers between the AVE frame and a desktop window frame.

By inputting to the PC 3 with a user input device such as a keyboard 14 or a mouse 15, the user 17 can move the virtual camera 71 such that the desktop 423 on the virtual computing appliance 421 is larger or smaller in the display device 264. The user 17 may also operate the desktop 423 on the virtual computing appliance 421 using input devices such as a keyboard 14 or a mouse 15.

This fourteenth embodiment of the avatar user interface system invention enables a person to shift between frames with low cognitive jolt. Advantages of this embodiment include: improved communication, better task efficiency, a more suitable interface for multi-tasking between verbal tasks and information tasks and higher usability.

Avatar agent sharing virtual computing appliance in AVE

It is a further purpose of this fourteenth embodiment that an avatar agent and a user may communicate in an AVE with a virtual computing appliance in it; the virtual computing appliance may be used by the avatar agent to communicate information to the user and by the user to communicate information to the avatar agent.

By way of disclosure of this fourteenth embodiment, a sample script is provided that might have been enacted between an avatar agent called Johan and a user using an AVE with a virtual laptop in it. The domain is the avatar agent giving professional advice to the user on risk management.

> 'Johan is the Advisor. He is a slightly old-fashioned 'Mad Professor' character, dressed in an old-style suit and bow tie. His half-glasses are at the end of his nose. He is seen seated at a desktop with a laptop facing towards the SME user. Behind him, through the glass panels of the meeting room, the user sees a huge, high-tech data centre which gives the impression of vast knowledge. Johan speaks

in a way that makes a bit of a caricature of himself. But he comes over as someone you would like and trust. In the Advice domain you can only talk to him; the mouse and keyboard are not used'

The opening shot sets the scene: camera at first person point of view of SME user; Johan's head, upper body visible plus data centre in background; virtual laptop partially visible on table orientated partly towards SME user

Johan "Hello, My name is Johan and I'm a most expert Risk Advisor. I'm a bit hard of hearing, it takes a while for me to work out what you say and I get easily confused. So please just answer my questions precisely and we'll be fine. If at any time I misunderstand you, just interrupt and say 'No, that's wrong'. Now, I'm a busy person so let's get on with it. First of all, which of these industries is your company working in?"

Camera pans and zooms to virtual laptop screen on which you see a list of industries.
Some seconds pass without the SME replying. Shot to Johan.

Johan "Come on, surely you know what industry your company is in. Just read out the closest industry."

SME "Woodworking" [Johan looks up in the air and thinks about this for a few seconds; this metaphor gives the intelligent agent time to plan a response]

Johan "I suppose you are consulting me because your workshop caught fire recently." [Johan laughs]
"Sorry, bad joke." [Slight pause, Johan leans forward]
"Right, I've looked at all the claims that we have had in the woodworking industry and these appear to be the risks in the Woodworking industry." [Shot to virtual laptop, showing list of risks]

"Please read each one. If it does not apply to you please say 'Not risk 7 or whatever the risk number is'. OK off you go and say finished when you are done"

SME [Several second pause]
"Not Risk 9"

Johan [Shot to Johan]
"No IT risk - you mean you don't have many PCs. What's the next risk you don't have?"

SME [Shot to laptop, short pause]
"Finished"

Johan [Shot to Johan]
"Great. This is your company's risk profile."
[Shot to laptop, risk severity/probability graph appears. Area under red line flashes]
"The risks under the red line are so small you don't want to worry about them. You run a successful business, you can probably stand small losses like those"
[Area above red line flashes]
"But you want to worry about those risks towards the top right"
[Worst risk flashes]
"Especially that one! Industrial accident risk. Could be nasty. Claims of more than a million Euro are not uncommon."

This embodiment is not limited to a single avatar agent, there may be a plurality of avatar agents interacting with the user. The virtual laptop is one example of a virtual computing appliance and other virtual computing appliance might be used in its stead such as virtual plasma screens. The user 17 may use input means to the AVE other than voice. Such means might include a keyboard 14 or mouse 15 which when used to create input, the input would appear directly on the virtual computing appliance display 422 visible on the display device 264.

FIFTEENTH EMBODIMENT

It is a purpose of this fifteenth embodiment that the avatar user interface system 261 comprises motion capture means and software director means to improve the sense of co-presence during a communication session.

Motion-tracking terminal

Figure 54 is a schematic of a motion-tracking terminal 265 of an avatar user interface system 261 including motion-tracking cameras 29 for a communication session. Three users 17 sit on chairs 174 around a table 172. At the end of the table 172 is a display device 264 with an AVE 420 displayed. The AVE 420 is displayed in such a way that the virtual table 51 in the AVE 420 appears to be a continuation of the physical table 172. Behind the virtual table 51 sit avatars 5 representing users 17 at other environmental locations 273. The AVE background behind the avatars 5 includes a virtual meeting room with windows 60, door 58, walls 55 and ceiling 56. Each user wears a microphone 12. There are loudspeakers 173 for outputting the voices of the participants that are not at that location. As disclosed in the Fourth Embodiment, sound is mixed.

Motion-tracking

The benefit of tracked avatar animation is to provide additional visual cues for facial expressions, head movements and hand gestures which contribute to natural face-to-face communication. Cameras 29 around the display device 264 capture the movements of participants. Images from the cameras 29 are processed in real-time to track facial animation, eye gaze, upper body movement and gestures. In a second step, the 2D tracked movements are mapped onto the 3D virtual environment and avatars of each person. In a third step, parameterised animation is generated. Video-based motion capture is used for non-invasive capture of face and body movements using a small number of cameras 29 surrounding the display screen 264. This motion capture augments the body and face movements of a participant's avatar animation where no motion capture input is available. A key innovation is the mapping of the captured movement to parameterised avatar motion models based on real movement to achieve realistic avatar animation that is robust to errors in the visual tracking.

Parameters control motion characteristics such as movement speed and size. The emotional content of the original movement is conveyed whilst avoiding artefacts due to errors in tracking. Adaptive background subtraction is used to separate foreground objects (people) from the background scene and avoid the requirement for highly structured backgrounds (blue-screen) or constant scene illumination.

Eye-Gaze Direction

Eye-contact is an essential visual cue in face-to-face communication. To establish eye-contact between a virtual avatar and real participant, eye gaze direction of all participants is reconstructed. In a virtual meeting it is critical to establish which participant each person is looking at in near real-time. To achieve this, key facial features are tracked for each participant using a statistical template of facial appearance for each individual based on their avatar model. This is used to robustly identify the location of the eyes at each time instant. The use of a model-based vision approach allows the three dimensional location of these facial features to be reconstructed. A dynamic eye-template which models the appearance of the eye with changes in viewing direction according to the iris location is then used to reconstruct the approximate viewing direction of the subject. Estimated gaze direction is used to identify if a participant is looking at the facial region of another real participant or avatar. Eye-contact is then established with the corresponding avatar. Avatar gaze-direction is animated to ensure correct eye-contact together with smooth transition of eye-contact between participants and with the background scene (ie the participant is not paying attention or looking at other documents).

Gesture Reconstruction

Established motion capture algorithms are used to reconstruct a subject's hand and head movement from the video streams. This approach utilises a real-time inverse kinematics engine to recover the approximate movement as estimates of joint angles. The reconstructed movement is mapped directly to the animated avatar using a dynamic filter to constrain the movement, impose joint angle limits and provide smooth animation. To achieve greater anima-realism, techniques for mapping the captured noisy movement into parameterised gestures

are used. A database of parameterised realistic gestures is established using conventional marker-based motion capture to construct models of common gestures and explicitly parameterise the intra-gesture variation. Statistical models based on learning from visual data identify the gesture class and map the gesture to the appropriate set of parameters. This model-based approach to gesture animation enables smooth and realistic gesture animation from noisy input data.

Facial Expression Recognition

A key visual cue in face-to-face visual communication is the secondary facial expression in conjunction with speech. A model-based methodology is adopted based on a highly sophisticated facial animation model. The facial animation model encodes parametric models of facial expression that express both the extent of movement and the temporal duration of the movement. Video analysis of facial expression using particle filters identifies key facial features corresponding to different facial expression. Statistical models of facial expression are learnt from labelled video sequence of multiple individuals. The learnt statistical models are used to identify the class of facial expression or combination of expressions. Finally detailed analysis of facial features is applied to identify the spatial and temporal parameters for a particular expression. The captured facial expression parameters are then used to augment the avatar facial movement synchronised with speech.

Multiple users

Although three users 17 are shown at the motion-tracking terminal 265 in Figure 54, this invention is operable for one or more users at each motion-tracking terminal 265. Depending on the detail design of a motion-tracking terminal 265, there are limits to the number of users 17 that can be motion tracked. One limitation comes when there are so many users close together that the motion tracking system cannot resolve which movements belong to which person. A second limitation is that of the computing power of the motion tracking system to follow a maximum number of users 17 simultaneously. A third limitation is from the number of chairs that can be fitted around the table 172. For large sessions, this motion-tracking terminal permits two or more

rows of chairs and for people to stand behind those sitting in the chairs. However, in this case most participants will not be motion tracked.

Setup

The input of users to the meeting consists of speech and motion. It is important that the captured speech and motion are attributed to the correct avatar on other user devices.

Speech is identified automatically by means of linking the identity of each microphone 12 with the avatar number 8 of the user 17. A person working in an organisation could have his identity card and his wireless microphone linked together. His microphone could be used for all voice input applications in the organisation such as fixed telephone, mobile telephone, paging, PC interaction and avatar user interface sessions. The organisation's database would link the person's identity, the microphone identity and the person's avatar number 8. This would be made available to the radio transceiver 170.

There are several ways of identifying the motion of a tracked person with the person's identity ie locating the person in the room. A low-technology way is a manual process using a seating plan. Chairs 174 are always in known positions and numbered: Chair 1, Chair 2 etc. In a manual setup process at the start of the session, a user 17 at each location identifies the avatar number of the person in each chair by means of direct input into the avatar user interface system 261, normally using keyboard 14 or mouse 15. This manual setup process works but relies on fixed chair positions.

A more flexible manual process is the interactive identification of each user 17. The motion-tracking terminal 265 knows who is present but not where they are located. In a simple procedure at the start of the session, the software director 80 asks each user in turn to wave both arms until the motion tracking system has located him. This enables people to move chairs around to suit the number of people present. One drawback of this method is that if people move around, the system might lose them eg if they leave the room to get something and then return. A drawback with manual processes is that

identification can take some time if there are a lot of people present and this wasted time costs money.

Ideally, the system should be automatic. There are several methods of achieving this. A first method is that wireless microphones are automatically tracked by triangulation of the signal between two or more receivers to estimate the location of the person in the room. These estimated locations are automatically mapped onto the motion-tracking system output to identify each moving person automatically.

In a second method, each microphone on the system has a visible, signal emitting light that is tracked by the cameras. The code of the signal emitting light is unique and associated with the identity of the person. The cameras map the light to the movement of the person to automatically identify the person. The advantage of tracking the microphones is that the microphone will always be within a small distance from the head/neck of a person.

None of these manual and automatic methods are perfect. Each has its own advantages and disadvantages. It is a purpose of this invention that a setup means be provided, either manual or automatic, for locating the position of each participant in the room. Automatic methods are better than manual methods, since they are more robust to movement during the session.

Terminal sizes

The motion tracking terminal 265 might be designed as a range of different sizes and to different price points. A large motion tracking terminal 265 might use the whole wall of a room as the display device 264. This might be achieved by the wall being a special opaque screen for rear transmission and the projector in an adjacent room projecting the AVE 420 onto the screen such that it is visible to the users 17 in the meeting room. The width of the table 172 could be more than 5 metres; the shape could be elliptical on one side and straight on the display side. Two rows of chairs 174 might be provided. A large number of cameras 29 could be situated to track a large number of participants 17 sitting in the chairs 174. Each

participant in the room could see each other participant. It might have a maximum capacity of more than 20 motion tracked people.

A medium-size motion tracking terminal 265 might use two plasma screens situated on the end of the table 172. It might have a maximum capacity of 7 motion tracked people.

A smaller motion tracking terminal 265 might use one monitor on the end of the table 172. It might have a maximum capacity of 3 motion tracked people. A motion tracking terminal 265 could be installed at each of the offices of an international organisation.

User devices and combinations

Many different types of user device may be used in an avatar user interface system 261. For multiple users at one location, a motion-tracking terminal 265 may be the optimal user device. For a user in his office a PC 3 may be the best device; this PC 3 may or may not have a webcam 29 to track the movements of the user 17. Whilst on the move, a user may use a mobile device such as a wireless Personal Digital Assistant (PDA) with telephone to participate in a communication session. Caves 350, exercise stations 414 and VR Headsets 366 are other types of user device that may be be used in an avatar user interface system 261.

Figure 55 is a block diagram of apparatus for an avatar user interface system 261 with multiple user devices. A session server 1, an avatar hosting server 4, an avatar agent hosting server 321, a motion-tracking terminal 265, a CAVE 350 and a PC 3 are connected together by a network 2.

At the lowest level of usage, an avatar user interface system 261 may be operable with a minimum of one user device and one user 17. In the case of one user 17, the user is probably communicating with an avatar intelligent agent software unit 320.

The highest quality usage for the best sense of co-presence is when all the users 17 are using motion tracking terminals 265.

This invention provides for the reality that users 17 may not all be at locations where there are motion tracking terminals 265 available and provides for users being connected via a variety of different user devices to one session.

SIXTEENTH EMBODIMENT

It is a purpose of this sixteenth embodiment that the display device 264 of the avatar user interface system 261 includes two or more projection means.

AVE and Presentation resolution

With a single display means such as a computer screen, if the virtual presentation screen 53 in the Avatar Virtual Environment (AVE) 420 is small, then a presentation slide containing words that is projected onto the virtual presentation screen 53 will be unreadable. A typical computer screen will have 1024 pixels across and this might also be the width of a large meeting room media window 50 showing an AVE 420. If the virtual presentation screen 53 is in proportion with the whole virtual meeting room, then it may only have 200 pixels width. This is not enough pixels for resolving the words on a presentation slide.

The human eye, has great resolving power and a person may read a poster on a wall, even if the poster is quite small and the person is not close to it. From the same position, the person can also take in the whole wall by 'zooming out'. A novel display apparatus in an avatar user interface system 261 is disclosed, which takes advantage of the capabilities of the human eye to view simultaneously, the AVE 420 and the presentation screen 53 at full resolution as if they were one environment.

Figure 56 is a schematic of a display device 264 consisting of a display screen 430, an AVE projector 431 and a Presentation projector 432. The meeting room media window 50 is projected by the AVE projector 431. The virtual presentation screen 53 is projected by the Presentation projector 432.

To avoid 'whiting out' the virtual presentation screen 53, the same area in the AVE is projected black with minimal light leaving the AVE

projector 431 to fall on the area of the presentation screen 53. In this way, the presentation benefits from the full contrast of the Presentation projector 432. Furthermore, the presentation appears brighter than the AVE, which is a strong parallel to a real presentation in a darkened real room, in which the presentation screen is usually the brightest element. Projection may be from the tabletop, from a ceiling attachment or in reverse from behind an opaque screen. The software director 80 on the PC 3 will generate two full-size displays: the AVE and the presentation; 3D graphics cards already on the market can drive two full-size displays. The display screen 430 may be any aspect ratio or it may be curved.

Dual projector unit

Figure 57 is a schematic of a display device 264 in which the AVE and Presentation projection means are combined into one physical unit 433. The AVE projection optics 434 has the normal controls available on a desktop projector such as focus and perhaps zoom. The axis 439 of the Presentation projection optics 435 may be altered such that it points anywhere within the AVE area 440 projected by the AVE projection lens 434. A slider control 436 can be moved by a user 17 to move the axis 439 from left to right. A slider control 437 can be moved by a user 17 to move the axis 439 up and down. A slider control 438 can be moved by a user 17 to zoom the Presentation area 441 in and out. The controls 436-438 may directly move the presentation projection optics 435, or they may drive motors that move the optics. In this way, the presentation area 441 can quickly be aligned to the right place in the AVE area 440 at the start of the session. During the session, it is important that, once set up, the software director 80, does not move the pixel position of the virtual presentation screen 30 in the AVE 420. Manual control of the position and size of the axis may be achieved by a number of other means such as the use of a remote control. A camera 221 built into the projector 443 that images the AVE area 440 could be used to locate the projected size/position of the Presentation area 441. A control loop could be constructed to set the presentation projector axis orientation/zoom automatically using software-driven motors driving the presentation projection optics 435. The control loop could be driven by the software director 80 from the PC 3 which could project reference images from both projectors

alternately that are imaged by the camera 221. It is a further purpose of this sixteenth embodiment that the projection means is provided with alignment means that can be either manual or automatic or both.

Powerful projected conference systems

Larger conference facilities may require two or more presentation screens within the overall AVE display. One presentation screen might show a video, a second a presentation slide and a third might show a head/shoulders shot of the avatar of the presenter. Each presentation screen might be driven by a different projector. Or, a plurality of virtual presentation screens might be arranged in the AVE such that they can be driven by one presentation projector 432. In this case, the resolution of each virtual presentation screen is half or less.

PCs are able to generate real-time 3D with more pixels than display projectors can project. Two or more AVE projectors 431 could be used in a tile formation to project a high-resolution AVE. Alignment means permit the projectors to be aligned to each other so that there are no gaps and no overlaps. The display screen 430 may be planar-rectangular, or it may be curved, or it may comprise a number of planes abutting at any angle. Different projectors might be located to project onto different planes or curves.

It is a further purpose of this sixteenth embodiment that any number of AVE projectors 431 and any number of Presentation projectors 432, whether integrated in units 433 of two or more projectors or not, may be used to display any number of virtual presentation screens within an AVE on a continuous display screen of any shape or combination of shapes.

Multi-density display device

Display devices available today usually have a single screen that is either illuminated within its unit (such as CRT monitors, LCD displays, plasma screens, opto-polymer displays) or comprises a separate screen illuminated by projection from another unit (front projector, rear projector). The scope of this sixteenth embodiment is not limited to projection devices, but includes single unit devices

with two or more areas of display of different pixel densities as measured by pixel row and column spacings in units of length.

Figure 58 is a schematic of a multi-density display device 451 comprising an area of low-density pixels 452 and an embedded area of high-density pixels 453. The multi-density display device 451 may be packaged in a single unit, which has the advantages of lower complexity, lower weight, lower manufacturing and lower installation costs for example. Or, it may be packaged as two or more units. The embedded high-density area 453 may insert into the low-density area 452 such that the join cannot be seen when the multi-density device is in use, or the join may be visible, but not in such a way that it impairs the usability of the device. The high-density area 453 could be situated anywhere in the low-density area 452. The high-density area 453 could be central, surrounded on all sides by low-density pixels 452, or it could be in an edge or at a corner or as a flap along a whole edge.

The main advantage of a multi-density display device 451 over a uniformly high-density device is that it will be lower cost to manufacture and require less electronics to drive. Most multi-density display devices 451 will only have double the number of pixels of a conventional display device, instead of possibly nine times for a typical application.

In general use, the multi-density display device 451 is operable such that a single image eg a photograph, can be displayed at uniformly low resolution across the entire device. There are advantages of pixel alignment such that the row and column density of the high-density area 453 is an integer factor of the low-density area 452. If the integer factor is 2 then there will be two rows of pixels in the high-density area for each row in the low-density area. The same applies for columns. This is shown in the magnified part of Figure 58. In this configuration, four high-density pixels 455 may be imaged to be equivalent to a single low-density pixel 454. A similar correspondence applies for other integer factors such as 3 or 4. It is also contemplated in this embodiment that there may be a different integer factor for columns than for rows and that there may be a real

factor such as 2.5 for rows and a different real factor such as 2.7 for columns. In the case of real factors, well-known image processing techniques may be used such that the display of a single image is not impaired. The display illumination intensity of a low-density pixel 454 may not be the same display illumination intensity as a high-density pixel 455. After a process of calibration, the image processing software will need to compensate for any difference in display illumination intensity in applications such as the display of a single photographic image.

In specialist use, such as in an avatar user interface system, the multi-density display device 451 may display an Avatar Virtual Environment 420 onto the low-density area 452 and a virtual presentation screen 53 onto some or all of the high-density area 453. In specialist use, the display illumination intensity of the low-density area 452 may be different from the display illumination intensity of the high-density area 453. In the case of displaying small text on the high-density area 453, it will be easier to read if it has a higher display illumination intensity.

Multi-density display devices 451 may be manufactured in a variety of ways using a variety of technologies such as liquid crystal, plasma and opto-polymers. Manufacturing processes will need to be developed for the production of multi-density display devices and this is not expected to be difficult for those skilled in the art.

It is a further purpose of this sixteenth embodiment that any number of low-density areas 452 and any number of high-density areas 453 may be combined in any way in a multi-density display devices 451.

Dual-projector / multi-density display device use

Dual-projection devices 433 and multi-density display devices 451 are useful in communication sessions involving both AVEs and detailed information displays. A key advantage is the combination of sense of presence and the ability to view detailed information such that the user has a feeling of being there. A range of devices 431, 432, 433, 451 may cover needs from one user in a small room to several thousands of users in a large conference room.

It is a further purpose of this sixteenth embodiment to disclose a process wherein a computing appliance means uses a display device comprising two projector means comprising the following steps:

- a first projector projects an avatar virtual environment;
- a second projector projects a presentation;

such that both projections respond to changes independently and at the frame rate being used.

SEVENTEENTH EMBODIMENT

It is a purpose of this seventeenth embodiment that the avatar user interface system 261 includes a directional microphone device 460.

Remote Presentations

As disclosed in this seventeenth embodiment, live presentations can be delivered by a remote presenter to a room with an audience using an avatar user interface system 261. Furthermore, live presentations can be delivered to a mixed audience consisting of an audience physically present in a room and a virtual audience simultaneously present at one or more other locations, connected by a network. During a presentation, the presenter's avatar can use media such as slide images projected onto a virtual screen.

In an interactive session with the audience, there are several problems. The first problem is that of gaze. It is normal for a lecturer to address the person in the audience who asked the question. But where is that person? In the second problem, that of mixed audiences, if the questioner is not in the same room as a viewer, then it will be beneficial for the viewer to see a virtual audience. A third problem is the probability of everyone in a large audience not having identifiable avatars and personal microphones.

Figure 59 is a schematic of an avatar user interface system 261 with a mixed audience of avatars 5 of virtual users at various locations and physical users 17 in an environmental location 273 which is a room containing the physical audience and a directional microphone device 460 that can not only record sound but also the direction from which the sound is coming. The directional microphone device 460 is

connected to a 'Room PC' 3 that is also connected to the room's display device 264 and a network 2. An avatar 5 labelled 'Virtual Presenter' represents a remote user 17 labelled 'Remote Presenter'. The remote presenter 17 is using a 'Presenter PC' 3 on the network 2. A physical user 17 labelled 'Questioner' asks a question. The voice 270 and its direction are picked up by the directional microphone device 460 that feeds the information to the PC 3. The software director 80 controls the gaze direction of the virtual presenter 5 to face the questioner 17 as the presenter 17 replies. The accuracy of the gaze direction of the virtual presenter 5 towards the questioner 17 can be improved by building a virtual model of the environmental location 273 including the positions of the display device 264 and the directional microphone device 460.

In large conference rooms, there are often a number of fixed microphones for use by the audience. The accuracy of the gaze direction can be further improved by (a) using the directional microphone device to identify which fixed microphone is being used and (b) use the known location of the fixed microphone in the virtual model of the environmental location to determine the gaze direction.

A 'Remote Questioner' 17 is visualised at the environmental location 273 as a 'Virtual Remote Questioner' 5 displayed on the display device 264. When the presenter responds to the remote questioner, the software director knows the positions of both the virtual presenter avatar 5 and the virtual remote questioner avatar 5 and can calculate the gaze direction. The physical members of the audience at the environmental location 273 see the remote presenter answering the remote questioner.

This embodiment is applicable to multiple remote presenters such as a presenter and a chairman or a panel of presenters. One or more of the presenters may be at the same environmental location 273. Any number of environmental locations 273 with two or more users 17 and any number of environmental locations 273 with one user 17 may be connected by a network 2 during a presentation. This embodiment is also applicable to the simple case of one remote presenter presenting

to one physical audience, in which case there is no virtual remote audience.

It is a purpose of this embodiment to provide means for a remote presentation using the avatar user interface system.

Prepared presentation

In a live remote presentation, the software director 80 has to determine movements for the virtual presenter avatar 5 in real-time. Many body and facial gestures are normally timed by skilled presenters to fit in with the beginning and end of sentences. This is not possible in real-time for the software director 80 because it does not know when a sentence is due to begin or end.

A remote presenter may pre-record his presentation using a microphone to record the words as he speaks them. The software director 80 can then be used to prepare a better visual avatar presentation than the live presentation. This preparation can be done automatically by the software director 80 or interactively with the presenter 17.

Figure 60 is a block diagram of an apparatus for presentation preparation. A presentation preparer 461 may be operated either automatically or interactively by a user 17 to output a prepared presentation 466. At any time later, the prepared presentation 466 may be played on a player 210. The presentation preparer 461 has a set of voice recordings 464 and any associated media elements 465 as the main input. Media elements might be slide images, animations, audio-video clips, 3D objects, avatar player scenes or any other type of media. A prepared presentation 466 is an example of an avatar player scene; it may be executed in a linear fashion by a player 210. A presentation may be prepared without media elements 465. A presentation might also be mimed without voice recordings 464.

In Automatic preparation, the software director 80 takes a series of voice recordings 464 that have been associated with presentation media elements 465 such as slide changes and automatically generates the complete presentation including but not limited to: movement, gestures, gaze and lipsync for avatars; lighting, prop and camera

animations. A library of presentation actions 462 and a presentation action generator 463 is used for preparing the avatar animation. A set of automatic presentation rules is built into the presentation preparer 461 which is a finite state machine.

In manual presentation preparation using the preparation preparer 461, the user 17 may select what animations should be used when. Manual preparation is based on manually editing event positions on a timeline.

There are several advantages of a prepared presentation: (a) nervous presenters can fully prepare their presentations with much less stress; (b) scope is removed for the presenter to poorly time his presentation and overrun the time slot; (c) the gestural quality and timing of the prepared presentation is higher; (d) unskilled presenters with poor body language need not be embarrassed.

It is a purpose of this embodiment to provide means for preparing a remote presentation using the avatar user interface system.

Presentation control

During the remote presentation, either a user 17 may control the mode of the software director 80 using mode selection buttons in the avatar user interface window 260, or the software director 80 may make a best guess at the mode. The rules applied to controlling the movement of the avatar of the presenter vary with mode. Modes include:

- Playing a prepared presentation
- Live presentation
- Question
- Answer
- Asking for questions
- Applause
- Background murmur between presentations

It is a purpose of this seventeenth embodiment to disclose a process wherein wherein directional microphone means and seating plan means are used comprising the following steps:

- a person speaks;
- a directional microphone means records the person's speech and the direction the speech is coming from;
- a software director uses the seating plan and the direction that the speech is coming from to generate avatar enactments such that displayed avatars can gaze in the direction of the speaker.

FURTHER MODIFICATIONS AND AMENDMENTS

Although the previous embodiments of the present invention have been described in which the personal computer 3 has been used for running the avatar user interface 160, it will be appreciated that a wide range of computing appliances 3 could be used. It will also be appreciated that any network may be used including the internet, a corporate intranet, an extranet, a virtual private network, wireless networks such as 3G, GSM, home wireless networks and direct connections such as ISDN or PSTN telephone. It will further be appreciated that when a user is referred to in this disclosure in the male form such as 'he', that this is an inconvenience of language and that the meaning is equally applicable to male and female users and that the use of this invention is not limited to males but may be used in an identical way by females.

CLAIMS

1. An apparatus for an avatar user interface system comprising:
- server means for serving the communication session;
- one or more computing appliance means;
- network means for joining said server means and said computing appliance means;
- avatar means for representing each user visually; and
- avatar user interface application means resident on each computing appliance means;
operable by one or more users.

2. Apparatus in accordance with claim 1 wherein said avatar means comprises an identity.

3. Apparatus in accordance with claim 2 wherein said identity means comprises an avatar number.

4. Apparatus in accordance with claim 3 wherein said avatar number means comprises an avatar hosting service identity number and an avatar identity number.

5. Apparatus in accordance with any of claims 2 to 4 wherein said identity means comprises a password.

6. Apparatus in accordance with any of claims 2 to 5 wherein said identity means comprises a display permission.

7. Apparatus in accordance with any of claims 2 to 6 wherein said identity means comprises biometric data.

8. Apparatus in accordance with any of claims 2 to 7 wherein said identity means comprises impersonation parameters.

9. Apparatus in accordance with claim 8 wherein said impersonation parameters means comprise action impersonation parameters.

10. Apparatus in accordance with claim 9 wherein said action impersonation parameters means is generated using an action impersonation parameter generator.

11. Apparatus in accordance with claim 8 wherein said impersonation parameters means comprise voice impersonation parameters.

12. Apparatus in accordance with any of claims 2 to 11 wherein said identity means comprises personal data.

13. Apparatus in accordance with any of claims 2 to 12 wherein said identity means comprises billing data.

14. Apparatus in accordance with any preceding claim wherein said avatar means comprises a 3D avatar.

15. Apparatus in accordance with claim 14 wherein said 3D avatar means comprises a parameter avatar.

16. Apparatus in accordance with claim 14 wherein said 3D avatar means comprises a photo-realistic avatar.

17. Apparatus in accordance with any of claims 1 to 13 wherein said avatar means comprises an animatable image avatar.

18. Apparatus in accordance with any of claims 1 to 13 wherein said avatar means comprises another avatar type.

19. Apparatus in accordance with any preceding claim wherein said avatar means is generated using an avatar generator editor.

20. Apparatus in accordance with claim 19 wherein said avatar generator editor means comprises a parameter avatar generator and a database for parameter avatars.

21. Apparatus in accordance with claim 19 wherein said avatar generator editor means comprises a photo-realistic avatar generator.

22. Apparatus in accordance with claim 21 wherein said photo-realistic avatar generator means comprises a booth.

23. Apparatus in accordance with claim 21 wherein said photo-realistic avatar generator means comprises a camera.

24. Apparatus in accordance with claim 21 wherein said photo-realistic avatar generator means comprises a service.

25. Apparatus in accordance with any preceding claim wherein said network means comprises an IP network.

26. Apparatus in accordance with any preceding claim wherein said network means comprises a plurality of networks.

27. Apparatus in accordance with claim 26 wherein said plurality of networks comprises at least one IP network and at least one telephone network.

28. Apparatus in accordance with claim 26 wherein said plurality of networks comprises at least one IP network; at least one telephone network and at least one mobile phone network.

29. Apparatus in accordance with any preceding claim wherein said avatar means is hosted by an avatar hosting server means.

30. Apparatus in accordance with claim 29 wherein said avatar hosting server means comprises memory means for storing avatars.

31. Apparatus in accordance with any of claims 29 or 30 wherein said avatar hosting server means comprises database means.

32. Apparatus in accordance with any of claims 29 to 31 wherein said avatar hosting server means comprises avatar hosting server management software means.

33. Apparatus in accordance with any of claims 29 to 32 wherein said avatar hosting server means comprises one or more avatar converter software means for converting avatar means from one format to another.

34. Apparatus in accordance with any of claims 4, 29 to 33 wherein said avatar hosting server means is identified by said avatar hosting service identity number and an avatar hosting registry server connected to said network stores location information as to the network location of said avatar hosting server means indexed to said avatar hosting service identity number operable such that the network location of said avatar hosting server means may be retrieved from said avatar hosting registry server means by provision of said avatar hosting service identity number.

35. Apparatus in accordance with any preceding claim wherein said server means for serving the communication session comprises session management software.

36. Apparatus in accordance with any preceding claim wherein said server means for serving the communication session comprises an event accumulator.

37. Apparatus in accordance with any preceding claim wherein said server means for serving the communication session comprises an audio mixer.

38. Apparatus in accordance with any preceding claim wherein said server means for serving the communication session comprises a text chat engine.

39. Apparatus in accordance with any preceding claim wherein said server means for serving the communication session comprises an e-mail engine.

40. Apparatus in accordance with any preceding claim wherein said server means for serving the communication session comprises a speech recognition engine.

41. Apparatus in accordance with any preceding claim wherein said server means for serving the communication session comprises a translation engine.

42. Apparatus in accordance with any preceding claim wherein said server means for serving the communication session comprises a text to speech engine.

43. Apparatus in accordance with any preceding claim wherein said server means for serving the communication session comprises a protocol converter.

44. Apparatus in accordance with any preceding claim further comprising an avatar agent hosting server.

45. Apparatus in accordance with claim 44 wherein said avatar agent hosting server comprises avatar agent hosting server management software.

46. Apparatus in accordance with claim 44 wherein said avatar agent hosting server comprises at least one intelligent agent software unit.

47. Apparatus in accordance with claim 46 wherein said intelligent agent software unit comprises artificial intelligence software and a knowledge base.

48. Apparatus in accordance with claim 46 wherein said intelligent agent software unit comprises avatar text to speech software and said voice impersonation parameters.

49. Apparatus in accordance with any of claims 44 to 48 wherein there is one computing appliance used by one user and one intelligent agent software unit hosted by one intelligent agent hosting server.

50. Apparatus in accordance with any preceding claim wherein said computing appliance means comprises display device means.

51. Apparatus in accordance with claim 50 wherein said display device means comprise two projector means projecting an avatar virtual environment projection and a presentation projection such that the presentation projection is significantly smaller than and lies within the boundary of the avatar virtual environment projection.

52. Apparatus in accordance with claim 51 wherein said two projectors means are provided in one physical unit.

53. Apparatus in accordance with claim 50 wherein said display device means comprises a multi-density display device with a high density area set into a low-density display.

54. Apparatus in accordance with any preceding claim wherein said computing appliance means comprises lip sync generation means.

55. Apparatus in accordance with any preceding claim wherein said computing appliance means comprises a headset comprising both speaker and microphone.

56. Apparatus in accordance with any preceding claim wherein said computing appliance means comprises at least one radio transceiver and an earpiece with microphone and speaker worn by a user for wireless conversation.

57. Apparatus in accordance with any preceding claim wherein said computing appliance means comprises at least one directional microphone such that it is possible to identify the direction from which the voice of the speaker is coming.

58. Apparatus in accordance with any preceding claim wherein said computing appliance means comprises an identity source reader.

59. Apparatus in accordance with any preceding claim wherein said computing appliance means comprises a biometric device.

60. Apparatus in accordance with any preceding claim wherein said computing appliance means comprises game interface equipment.

61. Apparatus in accordance with any preceding claim wherein said computing appliance means comprises a motion tracking terminal.

62. Apparatus in accordance with any preceding claim wherein said computing appliance means comprises an exercise station.

63. Apparatus in accordance with any preceding claim wherein said computing appliance means comprises a Cave.

64. Apparatus in accordance with claim 63 wherein said Cave means comprises a motion capture system.

65. Apparatus in accordance with any preceding claim wherein said avatar user interface application means comprises an avatar user interface window displayed on said display device means.

66. Apparatus in accordance with claim 65 wherein said avatar user interface window means comprises a session user interface window.

67. Apparatus in accordance with claim 66 wherein said session user interface window means comprises a meeting room media window controlled by a software director.

68. Apparatus in accordance with claim 65 wherein said avatar user interface window means comprises attendees functionality.

69. Apparatus in accordance with claim 65 wherein said avatar user interface window means comprises switchboard functionality

70. Apparatus in accordance with claim 65 wherein said avatar user interface window means comprises exhibitor functionality.

71. Apparatus in accordance with claim 65 wherein said avatar user interface window means comprises identity functionality.

72. Apparatus in accordance with any preceding claim wherein said avatar user interface application means comprises an avatar virtual environment displayed on said display device means.

73. Apparatus in accordance with claim 72 wherein said avatar virtual environment means comprises a virtual computing appliance.

74. Apparatus in accordance with any preceding claim further comprising a game hosting server.

75. Apparatus in accordance with any preceding claim further comprising a prepared presentation prepared using presentation preparer means.

76. Apparatus in accordance with claim 1 wherein there are the same number of said computing appliances and said users; each said computing appliance is used by one said user; each said computing appliance is at a unique physical location such that the display on said computing appliance would normally only be clearly visible to its user and not to any other user; and there is a minimum of two computing appliances and two users.

77. Apparatus in accordance with claim 1 wherein there are less said computing appliances than said users and at least one said computing appliance is shared by a plurality of said users in the same physical location as said computing appliance.

78. Apparatus in accordance with any preceding claim wherein no user views his own avatar on the computing appliance he is using.

79. Apparatus in accordance with claim 1 wherein said server means for serving the communication session is the same physical unit as one said computing appliance means such that said physical unit contains the functions of both said server means and said computing appliance means.

80. A method of communication between a plurality of users via an avatar user interface system comprising the steps of:

- joining a plurality of computing appliance means and a server means for serving the communications session to start a communication session by means of a network;
- viewing the avatars of the users involved in the communication session on the said plurality of computing appliance means;
- a user first communicating into a computing appliance;
- one or more users receiving the first communication on one or more other computing appliances;
- avatars enacting the first communication on said computing appliances;
- a user responding to the first communication in a second communication;
- one or more users receiving the second communication on one or more other computing appliances;
- avatars enacting the second communication on said computing appliances;
- continuing the exchange of communications until the session is finished; and
- terminating the joining of the computing appliance means and the server means for serving the communications session to terminate the communication session.

81. A method of communicating between at least one user and at least one avatar agent via an avatar user interface system comprising the steps of:

- joining one or more computing appliance means, an avatar agent hosting server means hosting one or more intelligent agent software units and a server means for serving the communications session to start a communication session by means of a network;
- viewing the avatars of the said avatar agents and said users involved in the communication session on the said computing appliance means;
- a user or an avatar agent first communicating;
- if there are one or more users who did not first communicate, then the one or more users who did not first communicate receive the first communication on one or more other computing appliances;

- avatars enacting the first communication on said computing appliances;
- if there are one or more avatar agents who did not first communicate, then the one or more avatar agents who did not first communicate receive the first communication;
- a user or an avatar agent responding to the first communication in a second communication;
- one or more users or one or more avatar agents receiving the second communication;
- if there are one or more avatars receiving the second communication, then avatars enact the second communication on said computing appliances;
- continuing the exchange of communications until the session is finished; and
- terminating the joining of the computing appliance means, the avatar agent hosting server means and the server means for serving the communications session to terminate the communication session.

82. A method in accordance with any of claims 80 or 81 wherein said viewing step each computing appliance means is viewed by one user.

83. A method in accordance with any of claims 80 or 81 wherein said viewing step at least one said computing appliance is viewed by a plurality of said users in the same physical location as said computing appliance.

84. A method in accordance with any of claims 80 or 81 wherein said viewing step each user cannot view his own avatar on the computing appliance he is using.

85. A method in accordance with any of claims 80 or 81 wherein said steps of communicating and receiving a communication take place in parallel with a small time delay that is acceptable to the user.

86. A method in accordance with any of claims 80 or 81 wherein said step of enacting a communication comprises both audio and visual output such that movements of said avatar observed visually including

lip movements in a speaking avatar are synchronised with the audio voice.

87. A method in accordance with any of claims 80 or 81 wherein said joining step further comprises means for identifying a user with an avatar.

88. A method in accordance with claim 87 wherein said viewing step further comprises means for receiving an avatar of a user at said computing appliance.

89. A method in accordance with claim 88 wherein said avatar is received from an avatar hosting service.

90. A method in accordance with claim 89 wherein said avatar is first converted by avatar converter software at said avatar hosting service from one format to another.

91. A method in accordance with any of claims 80 to 90 wherein said joining step further comprises the following steps:

- user providing an avatar number and password;
- said computing appliance sends said avatar number and said password to the network location of an avatar hosting service;
- avatar hosting server management software on said avatar hosting service checks a database to verify that said avatar number and said password are valid;
- if said avatar number and said password are valid, then avatar hosting server management software on said avatar hosting service sends said avatar to said computing appliance.

92. A method in accordance with any of claims 80 to 90 wherein said joining step, an avatar number comprises an avatar hosting service identity number and an avatar identity number, and said joining step further comprises the following steps:

- user providing an avatar number and password;
- said computing appliance sends an avatar hosting service identity number to an avatar hosting registry server;

- said avatar hosting registry server sends to said computing appliance the network location of the avatar hosting service corresponding to said avatar hosting service identity number;
- said computing appliance sends said avatar number and said password to the network location of said avatar hosting service;
- avatar hosting server management software on said avatar hosting service checks a database to verify that said avatar number and said password are valid;
- if said avatar number and said password are valid, then avatar hosting server management software on said avatar hosting service sends said avatar to said computing appliance.

93. A method in accordance with claim 80 wherein said enacting step a viewing user at a computing appliance sees an avatar virtual environment with avatars of other users at other computing appliances that are photo-realistic and that move anima-realistically, that substantially gives said viewing user the impression of the other users being together in one virtual location.

94. A method in accordance with claim 81 wherein said enacting step a viewing user at a computing appliance sees an avatar virtual environment, with any avatars of other users at other computing appliances and any avatars of avatar agents, that are photo-realistic and that move anima-realistically, that substantially gives said viewing user the impression of any other users and any avatar agents being together in one virtual location.

95. A method in accordance with any of claims 80 to 91 wherein said enacting steps, software director means drive avatar engine means to generate said enactment.

96. A method in accordance with claim 87 wherein said joining step, said avatar means comprises an identity that further comprises a display permission, and said joining step further comprises means for checking that said display permission permits display of said avatar on computing appliance means.

97. A method in accordance with claims 80 or 81 wherein said joining step, said avatar means comprises an identity and said joining step further comprises the following steps:

- said person providing an identity source that is read by an identity source reader;
- retrieving the avatar whose identity matches the identity in said identity source;
- displaying said avatar;
- a security user visually comparing said avatar with said person.

98. A method in accordance with claims 80 or 81 wherein said joining step, said avatar means comprises an identity that further comprises avatar biometric data, and said joining step further comprises the following steps:

- said person providing an identity source that is read by an identity source reader;
- retrieving the avatar whose identity matches the identity in said identity source;
- extracting said avatar biometric data from said avatar;
- a biometric device scanning part of said person to provide scanned biometric data;
- comparing said scanned biometric data with said avatar biometric data;
- if said scanned biometric data does not match said avatar biometric data then alerting the security user;
- displaying said avatar to said alerted security user;
- said alerted security user visually comparing said avatar with said person.

99. A method in accordance with any of claims 80 or 81 wherein sound passes through the avatar user interface system comprising the following steps:

- a microphone means records sound from a user of a computing appliance means as said user speaks;

- a lip synchronisation generator means on said computing appliance means processes said sound to provide a combined audio and geometric position stream;
- the computing appliance means streams said combined audio and geometric position stream over the network to an audio mixer;
- said audio mixer mixes said combined audio and geometric position stream with any other combined audio and geometric position streams to produce a specific mixed audio and geometric position stream for each computing appliance;
- said audio mixer sends each computing appliance its specific mixed audio and geometric position stream;
- said computing appliance plays said specific mixed audio and geometric position stream to its user via a loudspeaker means.

100. A method in accordance with claim 99 wherein said lip synchronisation generator process comprises a process performed at regular intervals on a digital audio stream flowing into a buffer of the following steps:

- the contents of the buffer are copied and then the buffer is emptied;
- a discrete fourier transform is performed on the copied contents of the buffer and a spectrum is output;
- one or more analysers analyse the output spectrum and each analyser outputs a value representing a geometric position of a part of a talking head.

101. A method in accordance with claim 100 wherein the sequence of audio spectrums is combined with the sequence of geometric positions for transmission over the network.

102. A method in accordance with claim 101 wherein there are compression and decompression steps.

103. A method in accordance with claim 95 wherein said software generator uses personal action impersonation parameters defined for an avatar to generate animations for said avatar such that said avatar moves recognisably like the person it represents.

104. A method in accordance with claim 95 wherein said software generator uses generic action impersonation parameters defined for a communication context such that avatars move in ways believable within that communication context.

105. A method in accordance with claim 104 wherein said generic action impersonation parameters are defined for said communication context comprising the following steps:

- recording a corpus of videos of said communication context;
- processing said corpus by a trained person along a timeline to produce an annotated timeline with actions of each communication context participant related to a number of parameters;
- analysing said annotated timeline by a trained person to produce a type definition of each action impersonation parameter and a set of rules that can be incorporated into a finite state machine for said communication context.

106. A method in accordance with claim 103 wherein said personal action impersonation parameters for a particular person are generated using an action impersonation generator/editor means involving manual input by a user comprising the following steps:

- in the first step, said user makes selections from a number of sets of generic action impersonation parameters at a high level;
- in the second step, said user edits said selections at a lower level;

wherein said second step is optional and said user may or may not be the person for whom the personal action impersonation parameters are generated.

107. A method in accordance with claim 103 wherein said personal action impersonation parameters for a particular person are generated automatically using an action impersonation generator/editor means comprising the following steps:

- in the first step, video recordings are made of said person carrying out a number of defined actions;
- in the second step, the action impersonation generator/editor

automatically analyses the video recordings to generate a set of personal action impersonation parameters.

108. A method in accordance with claim 95 wherein said software director uses voice impersonation parameters defined for an avatar to generate speech from text using text to speech engine means for said avatar such that said avatar speaks recognisably like the person it represents comprising the following steps:

- intelligent agent software unit means generates said text;
- text to speech engine means converts said text to speech;
- said speech is played on said computing appliance.

109. A method in accordance with claim 108 wherein said voice impersonation parameters are defined for said avatar of a particular person comprising the following steps:

- recording said person speaking predefined text;
- processing said recording using impersonation parameter generation software;
- said impersonation parameter generation software outputting said voice impersonation parameters for that person;
- storing said voice impersonation parameters in said avatar.

110. A method in accordance with claim 108 wherein the person who is being impersonated is known to the user such that the avatar impersonating said person speaks and moves recognisably like said person.

111. A method in accordance with any of claims 80 to 109 wherein said avatar means comprises a 3D avatar.

112. A method in accordance with any of claims 80 to 109 wherein said avatar means comprises a parameter avatar.

113. A method in accordance with any of claims 80 to 109 wherein said avatar means comprises an animatable image avatar.

114. A method in accordance with any of claims 80 to 81 wherein said avatar means is first generated using an avatar generator editor.

115. A method in accordance with claim 114 wherein said avatar generator editor means comprises a booth.

116. A method in accordance with claim 114 wherein said avatar generator editor means comprises a camera.

117. A method in accordance with claim 114 wherein said avatar generator editor means comprises a service.

118. A method in accordance with claim 81 wherein after a voice communication by a user, a speech recognition engine means processes the voice communication comprising the following steps:

- a user generates a voice communication by speaking;
- a speech recognition means processes the voice communication and outputs text;
- the text is sent to any intelligent agent software units involved in the session.

119. A method in accordance with claim 118 wherein a user speaks in a first language and an intelligent agent software unit operates in a second language such that text is translated by translation engine means comprising the following steps:

- a user generates a voice communication by speaking in a first language;
- a speech recognition means that operates in said first language processes the voice communication in said first language and outputs text in said first language;
- the text in said first language is translated by translation engine means into text in a second language;
- text in said second language is sent to any intelligent agent software units involved in the session capable of processing text in said second language.

120. A method in accordance with any of claims 95 to 118 wherein a user understands a first language and an intelligent agent software unit operates in a second language such that text is translated by translation engine means comprising the following steps:

- an intelligent agent software units generates text in a first language;
- the text in said first language is translated by translation engine means into text in a second language;
- text to speech engine means converts said text in said second language to speech in said second language;
- said speech in said second language is played to said user using loudspeaker means.

121. A method in accordance with any of claims 80 to 120 wherein said computing appliance means uses a display device comprising two projector means comprising the following steps:

- a first projector projects an avatar virtual environment;
- a second projector projects a presentation;

such that both projections respond to changes independently and at the frame rate being used.

122. A method in accordance with claim 95 wherein directional microphone means and seating plan means are used comprising the following steps:

- a person speaks;
- said directional microphone means records said person's speech and the direction said speech is coming from;
- said software director uses said seating plan and said direction that said speech is coming from to generate avatar enactments such that displayed avatars can gaze in the direction of the speaker.

123. A method in accordance with any of claims 80 or 81 wherein users communicate whilst exercising on exercise station means comprising the following steps:

- a first user using a first exercise station means;
- a second user using a second exercise station means;
- said first user viewing the avatar of said second user using a virtual exercise station;

- said second user viewing the avatar of said first user using a virtual exercise station;
- said first and second users communicating by voice;
- optionally said first and second users viewing performance data generated by said first and second exercise station means;
- optionally any user being able to see if the other user has stopped exercising.

124. A method in accordance with any of claims 80 or 81 wherein users are present in Cave means with motion capture systems means comprising the following steps:

- said motion capture system means records movements of a first user in a first Cave means;
- said recorded movements are sent with acceptable lag from said first Cave means to a second Cave means;
- an avatar of said first user is displayed in said second Cave means such that the movements of said avatar duplicate the movements of said user in space;
- a second user wearing shutter glasses or similar immersive 3D viewing means in said second Cave means views the movements of the avatar of said first user as if said first user were physically in said second Cave with said second user.

125. A method in accordance with any of claims 80 or 81 wherein users communicate in virtual exhibition means comprising the following steps:

- a user navigates in a virtual exhibition stand of a company;
- said user views and interacts with virtual objects representing products;
- optionally said user communicates remotely with a real sales representative;
- optionally said user communicates with an intelligent agent avatar;
- optionally said user views presentations;
- optionally said user buys said product.

126. A method in accordance with any of claims 80 or 81 wherein users communicate in an interactive game hosted on a game hosting server comprising the following steps:

- a first user interacts with the game, navigates around the 3D game scene and views the avatar of a second user;
- said second user interacts with said game, navigates around said 3D game scene and views the avatar of said first user;
- said first user communicates by speaking;
- said second user hears said first user and views the avatar of said first user in lip synchronisation with said first user's speech;
- said second user communicates by speaking;
- said first user hears said second user and views the avatar of said second user in lip synchronisation with said second user's speech.

127. A method in accordance with any of claims 80 or 81 wherein a remote presenting user presents a presentation remotely comprising the following steps:

- said remote presenting user starts a prepared presentation;
- remote audience users watch the avatar of said remote presenting user perform said prepared presentation;
- present audience users present physically together in a theatre watch a projection of the avatar of said remote presenting user perform said prepared presentation;
- said prepared presentation ends;
- a remote audience user asks a question;
- said remote presenting user views the avatar of said remote audience user asking the question from amongst a single virtual audience and said avatar of said remote audience user gazes at said remote presenting user;
- said present audience users view the avatar of said remote audience user asking the question from amongst a single virtual audience around the avatar of said remote presenting user and said avatar of said remote audience user gazes at said avatar of said remote presenting user.

128. A method in accordance with claim 127 wherein a prior step, said remote presenting user prepares a presentation using presentation prepare means.

Fig 1

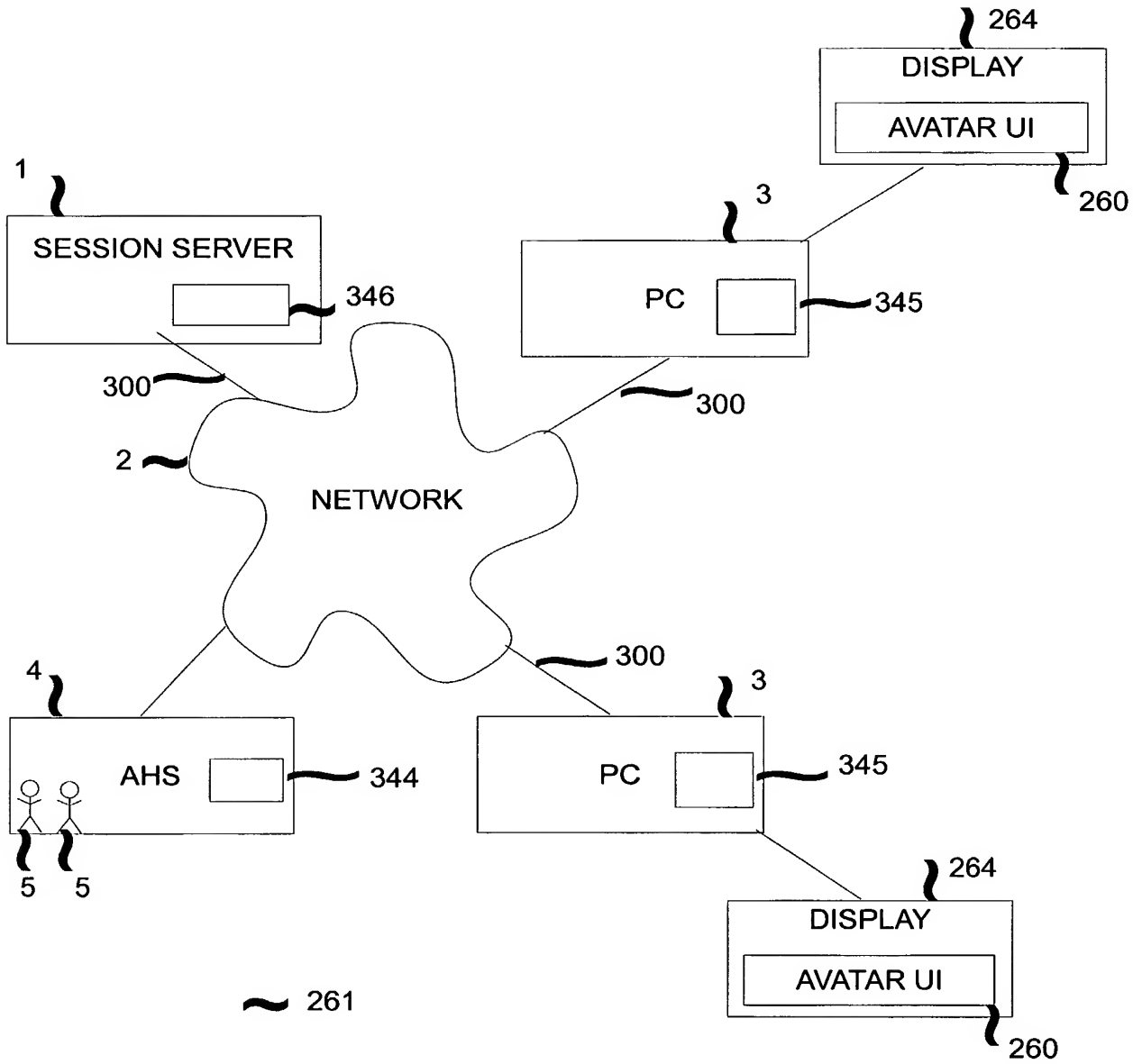

264
DISPLAY
AVATAR UI
260
1
SESSION SERVER
346
3
PC
345
300
300
2
NETWORK
300
4
AHS
344
3
PC
345
5
5
264
DISPLAY
AVATAR UI
260
261

Fig 2

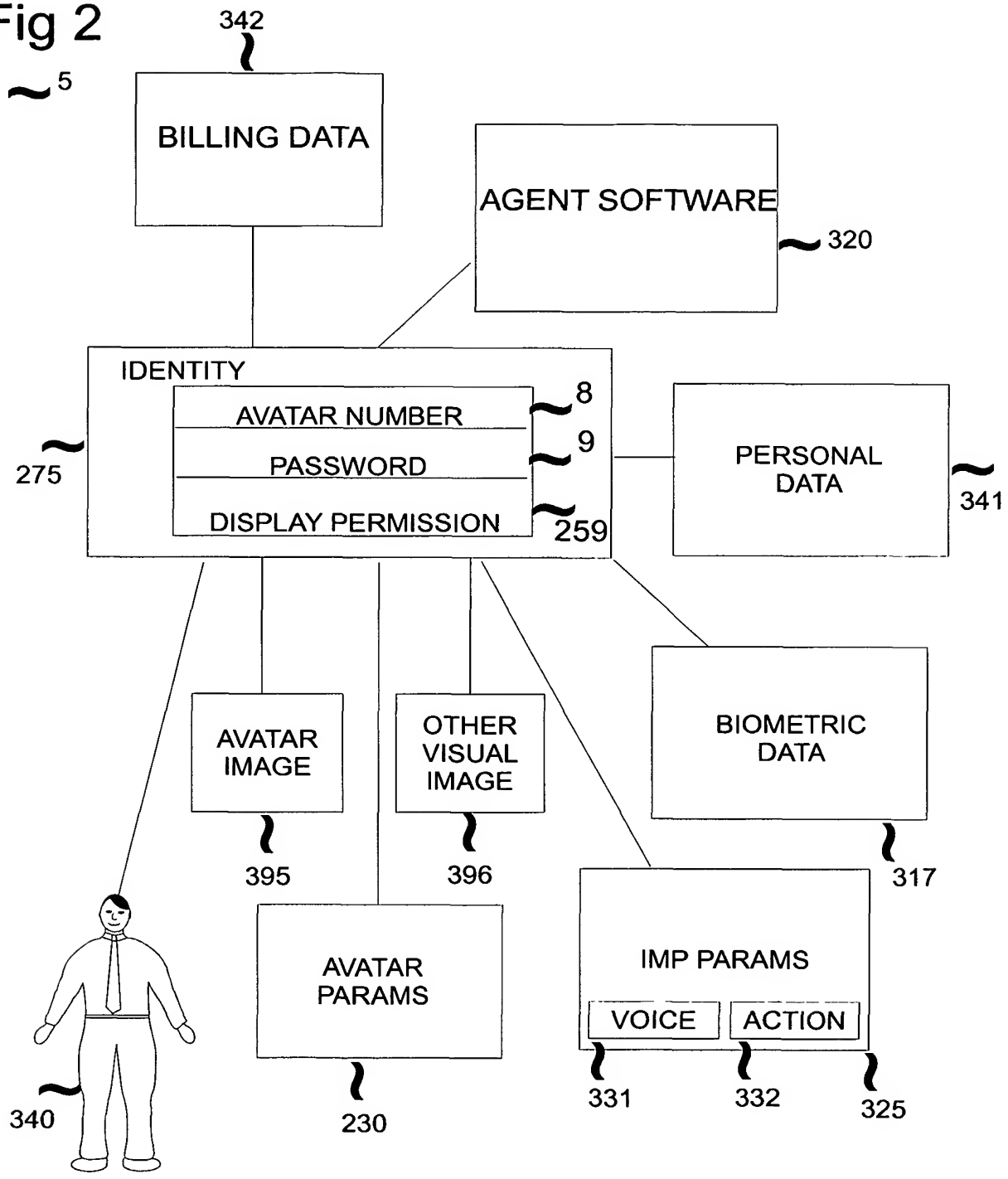

342
5
BILLING DATA
AGENT SOFTWARE
320
IDENTITY
AVATAR NUMBER
8
PASSWORD
9
DISPLAY PERMISSION
259
275
PERSONAL DATA
341
AVATAR IMAGE
OTHER VISUAL IMAGE
BIOMETRIC DATA
395
396
317
AVATAR PARAMS
IMP PARAMS
VOICE
ACTION
340
230
331
332
325

Fig 3

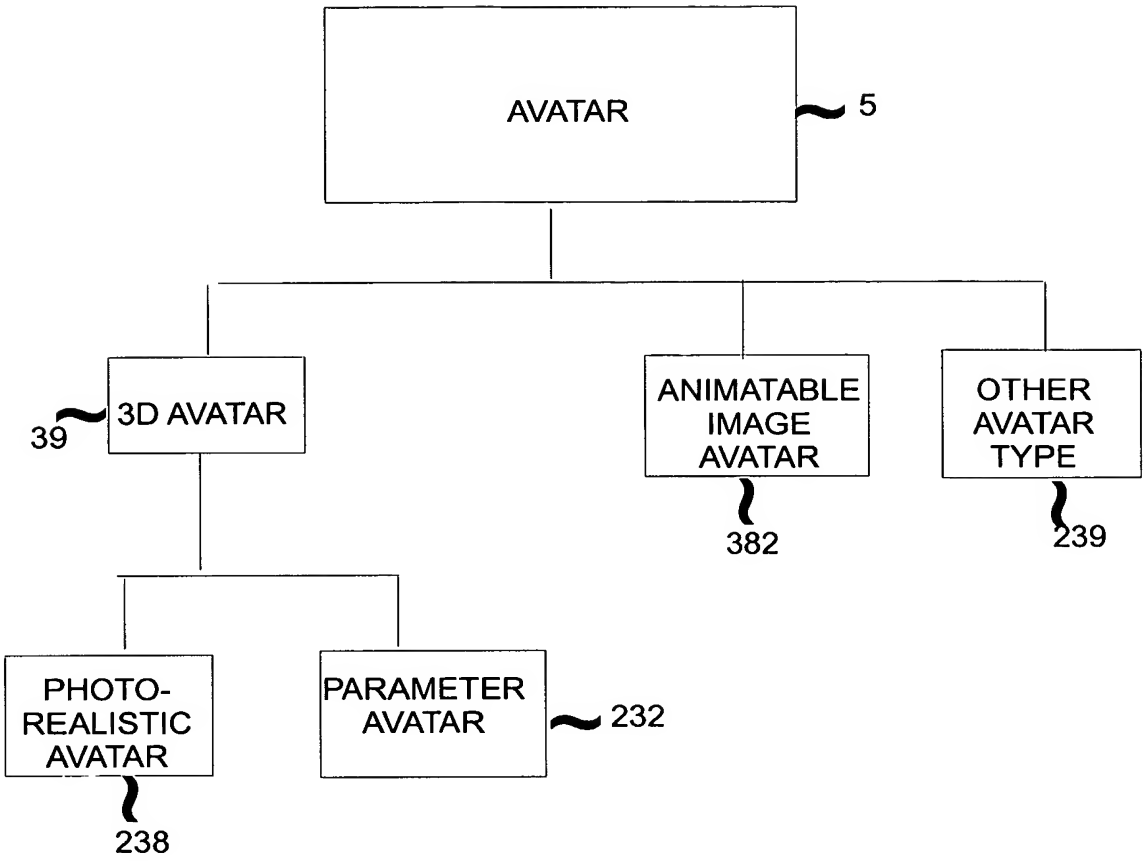

AVATAR
5
3D AVATAR
39
ANIMATABLE IMAGE AVATAR
382
OTHER AVATAR TYPE
239
PHOTO-REALISTIC AVATAR
238
PARAMETER AVATAR
232

Fig 5

Fig 4

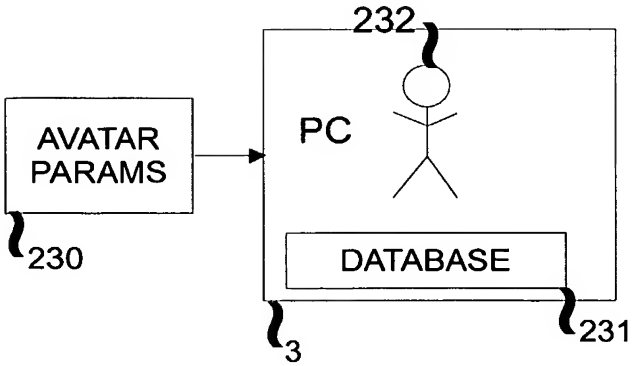


Fig 6

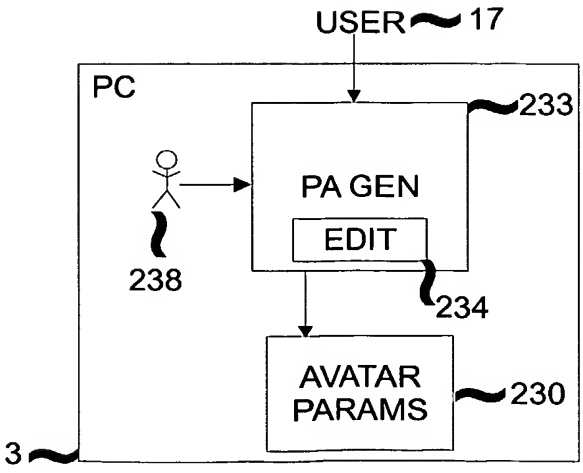


Fig 7

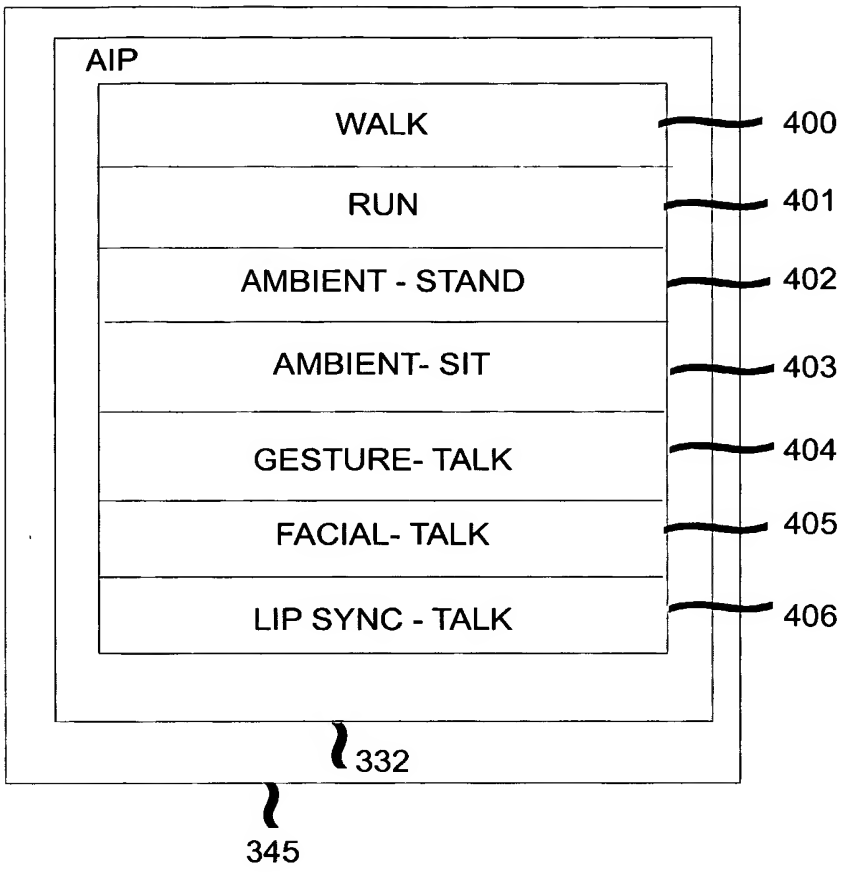

AIP
WALK
RUN
AMBIENT - STAND
AMBIENT- SIT
GESTURE- TALK
FACIAL- TALK
LIP SYNC - TALK
400
401
402
403
404
405
406
332
345

Fig 7a

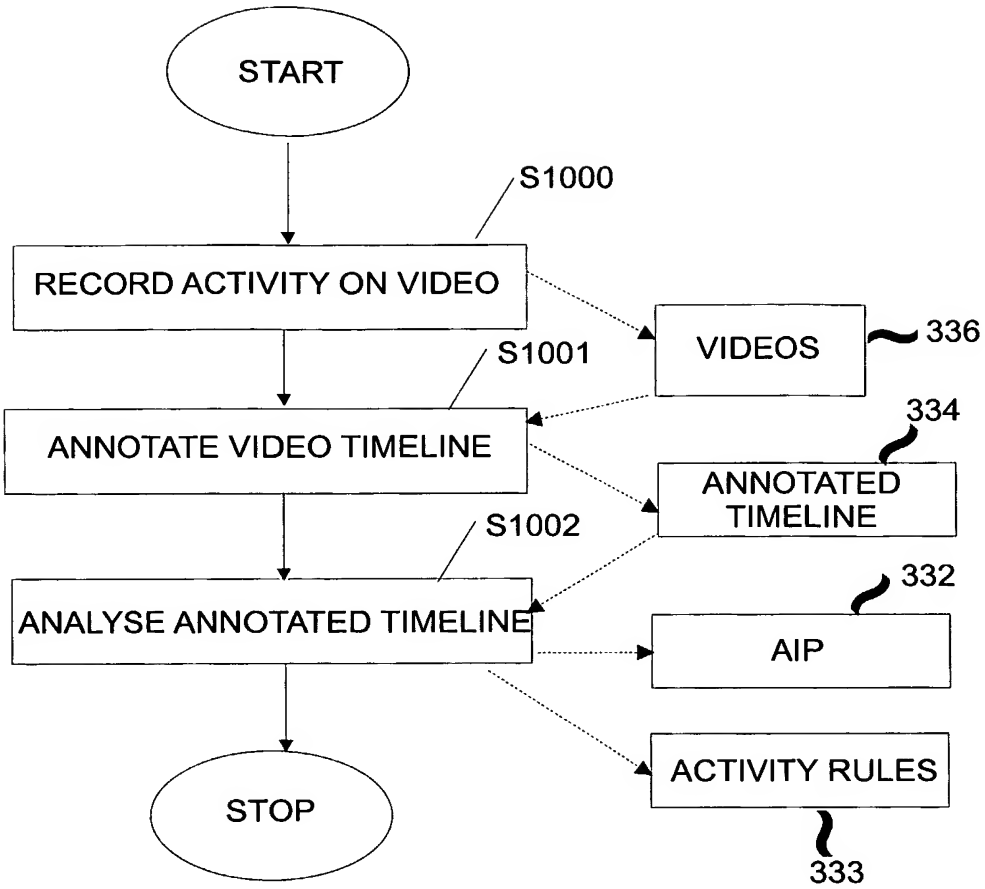

START
S1000
RECORD ACTIVITY ON VIDEO
VIDEOS
336
S1001
ANNOTATE VIDEO TIMELINE
334
ANNOTATED TIMELINE
S1002
ANALYSE ANNOTATED TIMELINE
332
AIP
ACTIVITY RULES
333
STOP

Fig 8

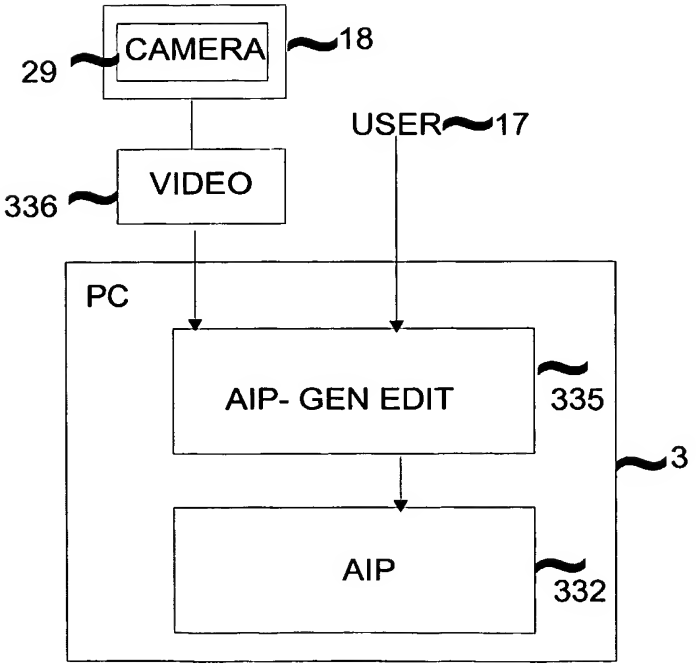

29
CAMERA
18
USER
17
336
VIDEO
PC
AIP- GEN EDIT
335
3
AIP
332

Fig 9

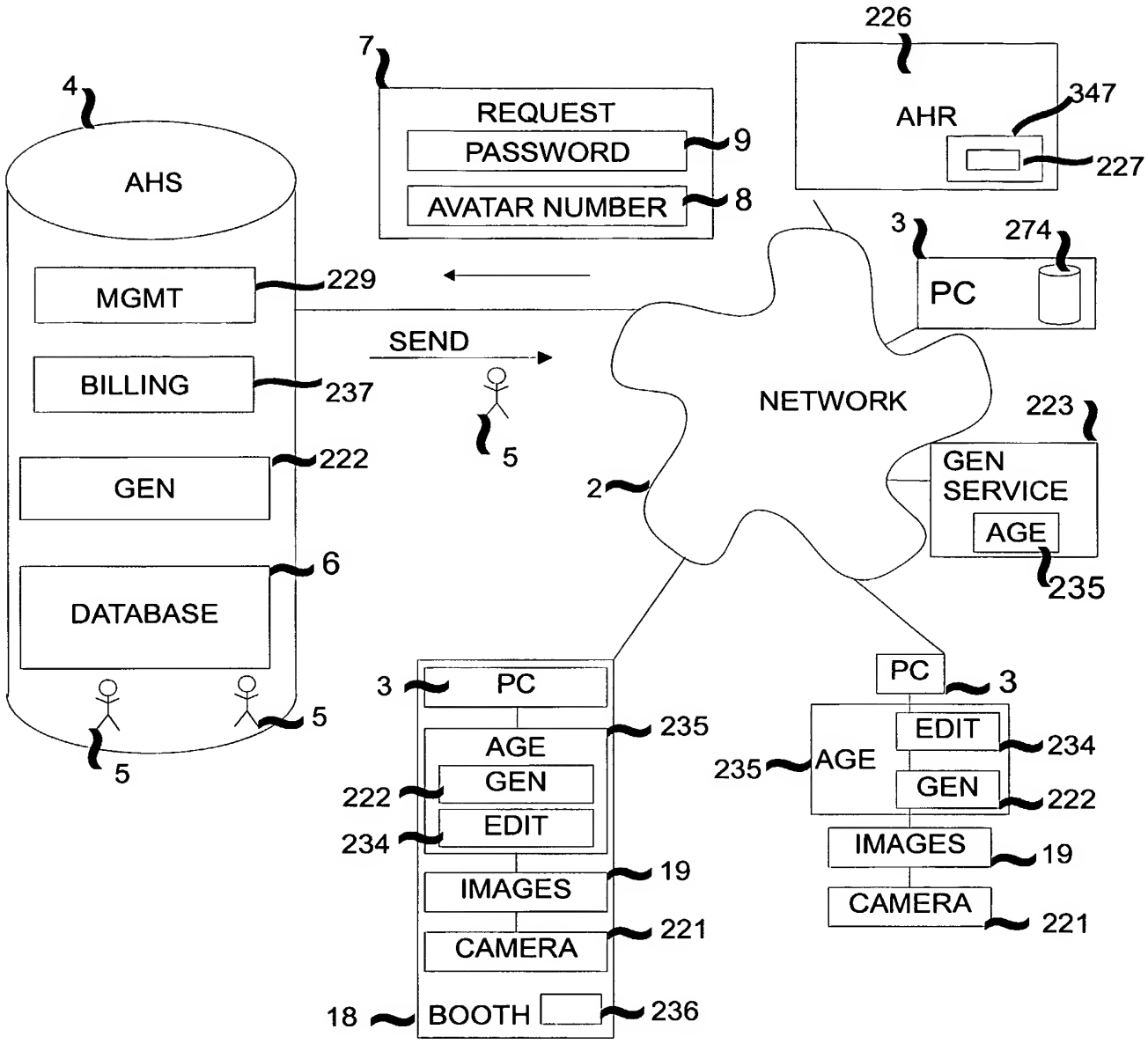

4
AHS
MGMT
229
BILLING
237
GEN
222
DATABASE
6
5
5
7
REQUEST
PASSWORD
9
AVATAR NUMBER
8
SEND
5
226
AHR
347
227
3
274
PC
NETWORK
2
223
GEN
SERVICE
AGE
235
3
PC
235
AGE
222
GEN
234
EDIT
19
IMAGES
221
CAMERA
18
BOOTH
236
PC
3
EDIT
234
235
AGE
GEN
222
IMAGES
19
CAMERA
221

Fig 10

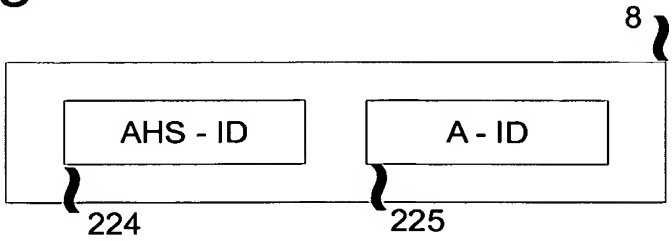

8
AHS - ID
A - ID
224
225

Fig 11

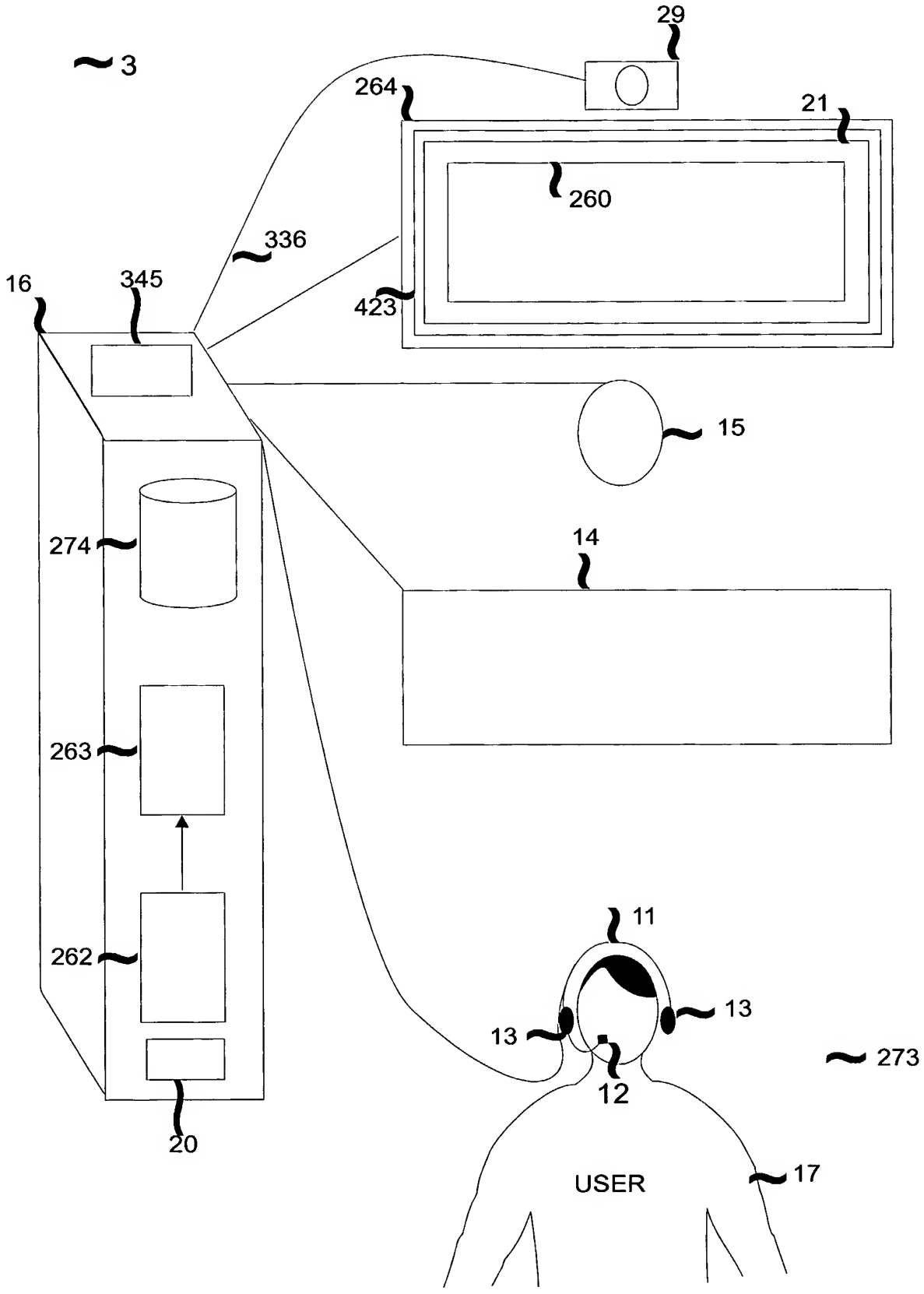

3
29
264
21
260
336
16
345
423
15
274
14
263
262
11
13
13
273
12
20
USER
17

Fig 12

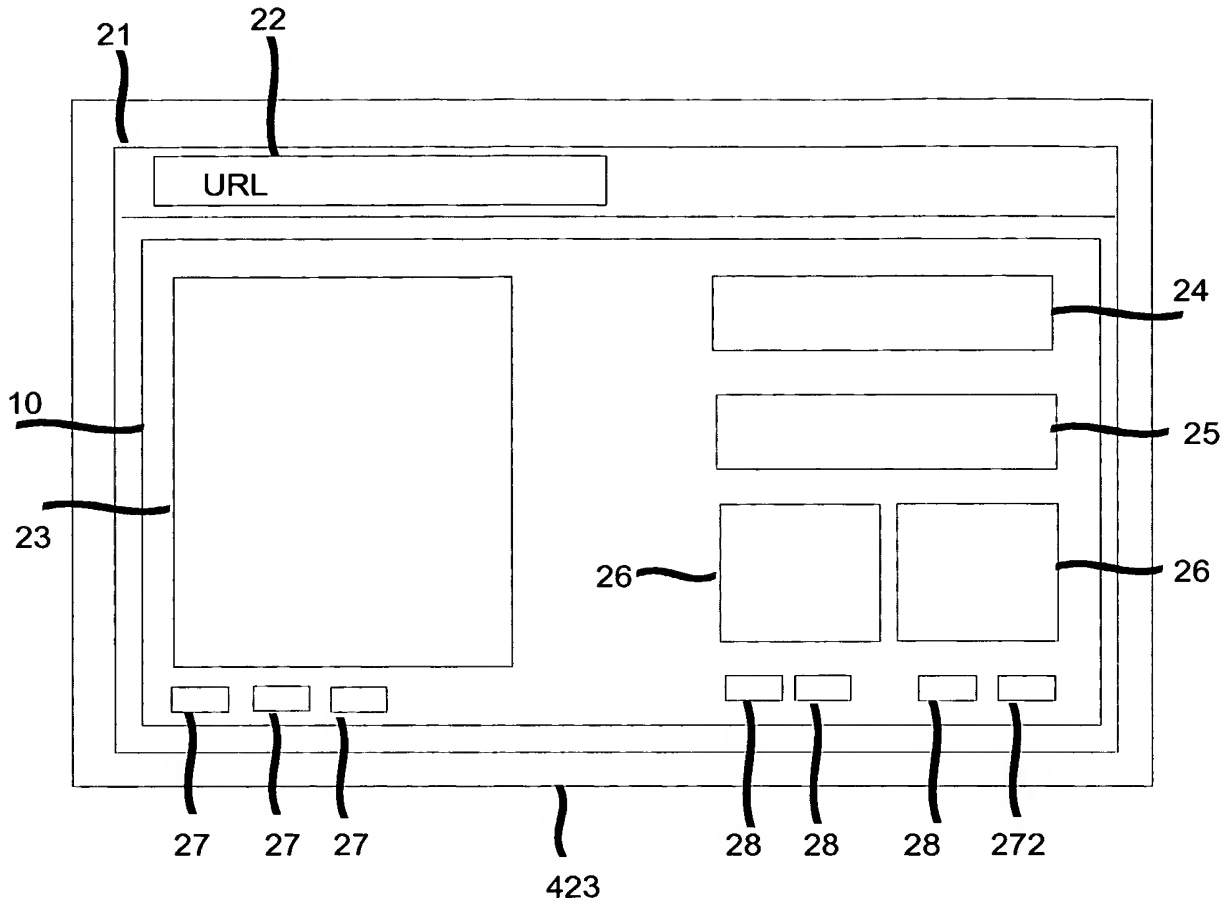


Fig 13

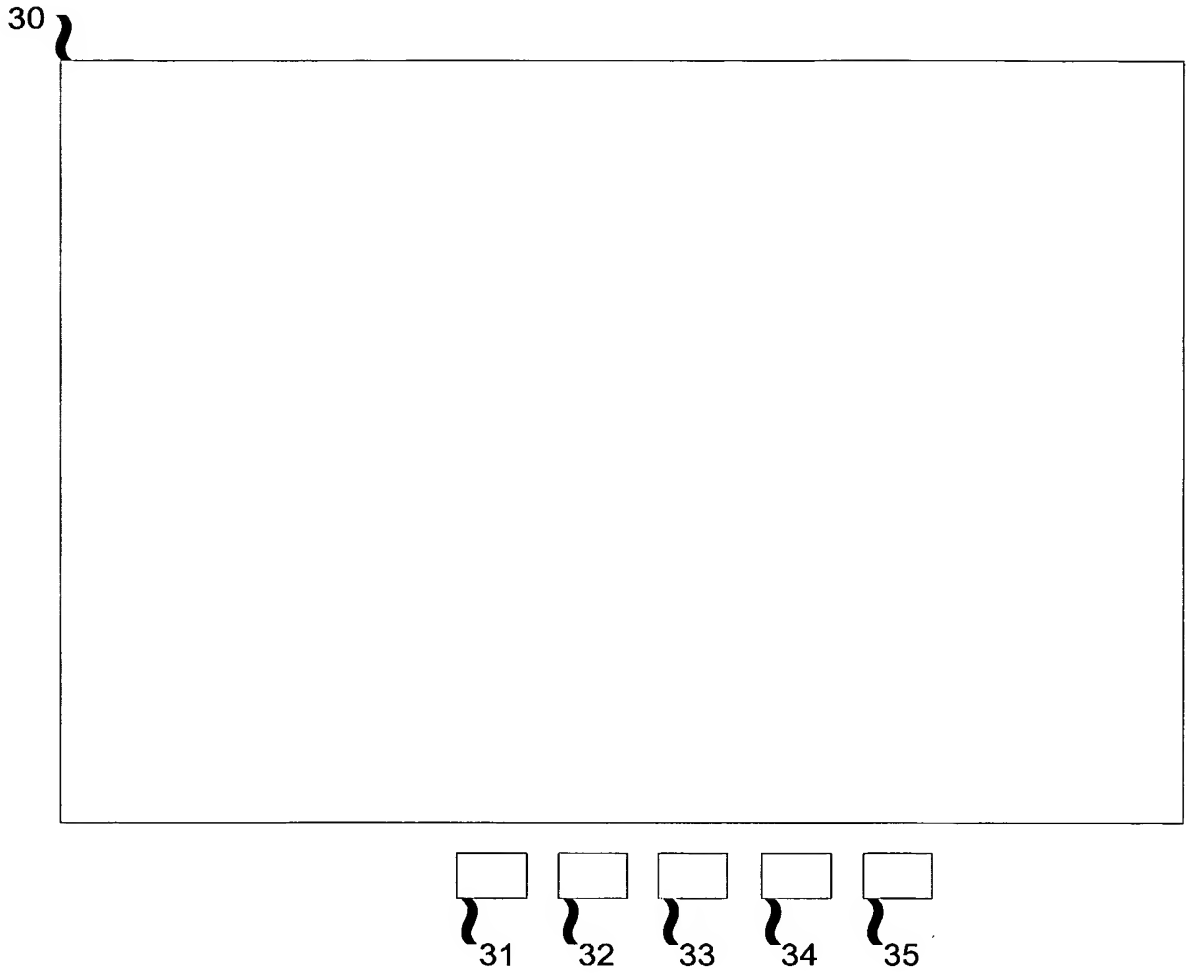


Fig 14

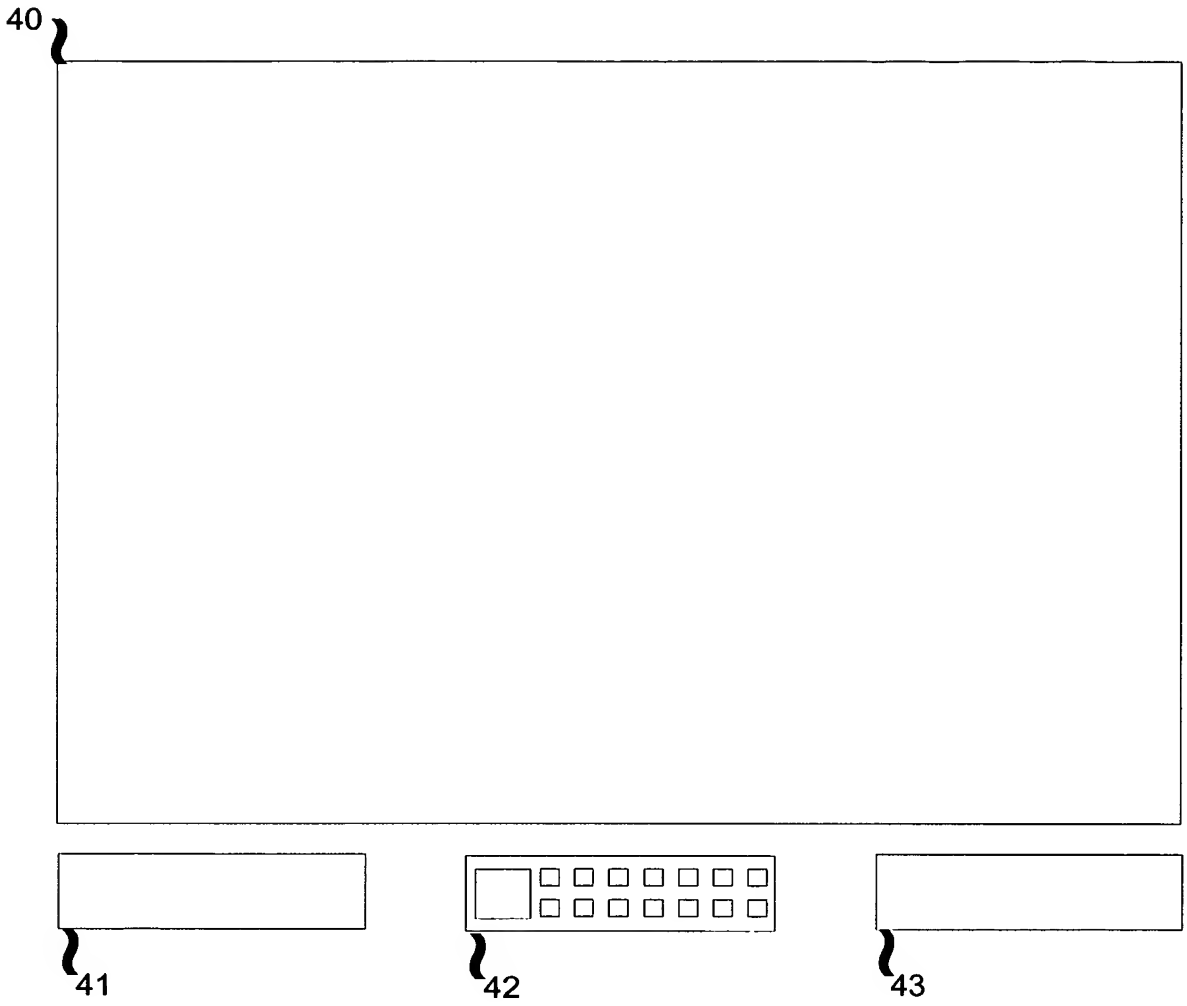


Fig 15

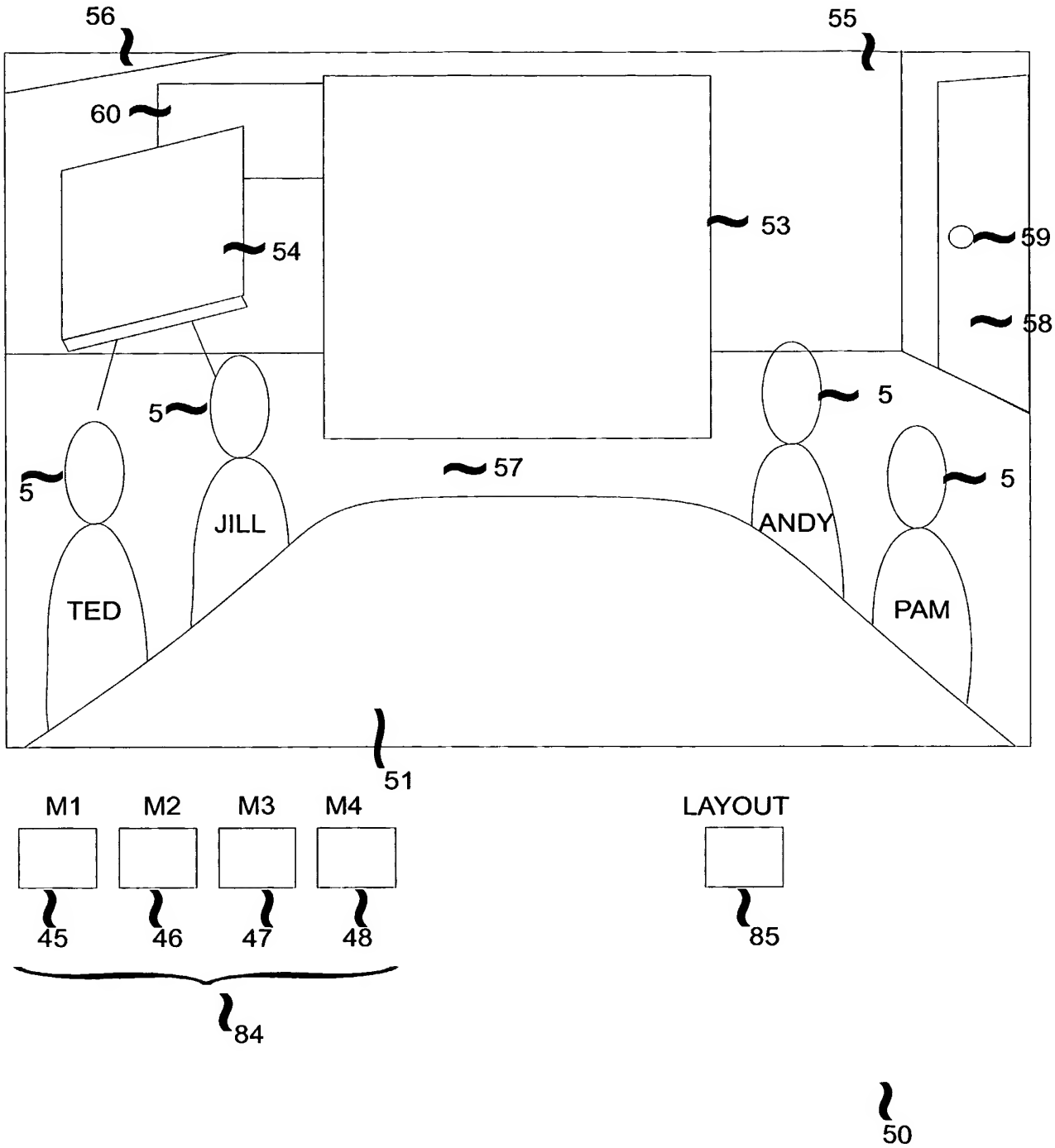

56
55
60
53
54
59
58
5
5
5
5
57
JILL
ANDY
TED
PAM
51
M1
M2
M3
M4
LAYOUT
45
46
47
48
85
84
50

Fig 16a Fig 16b Fig 16c Fig 16d

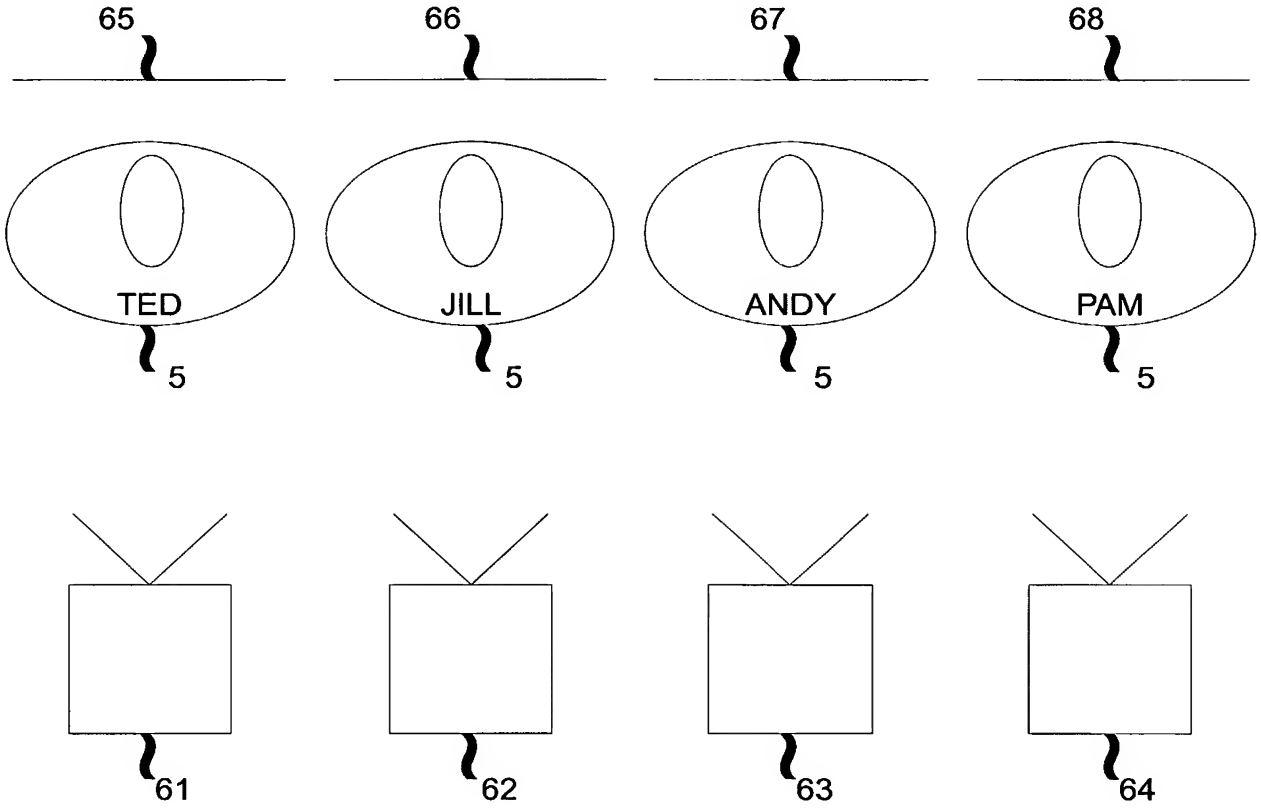
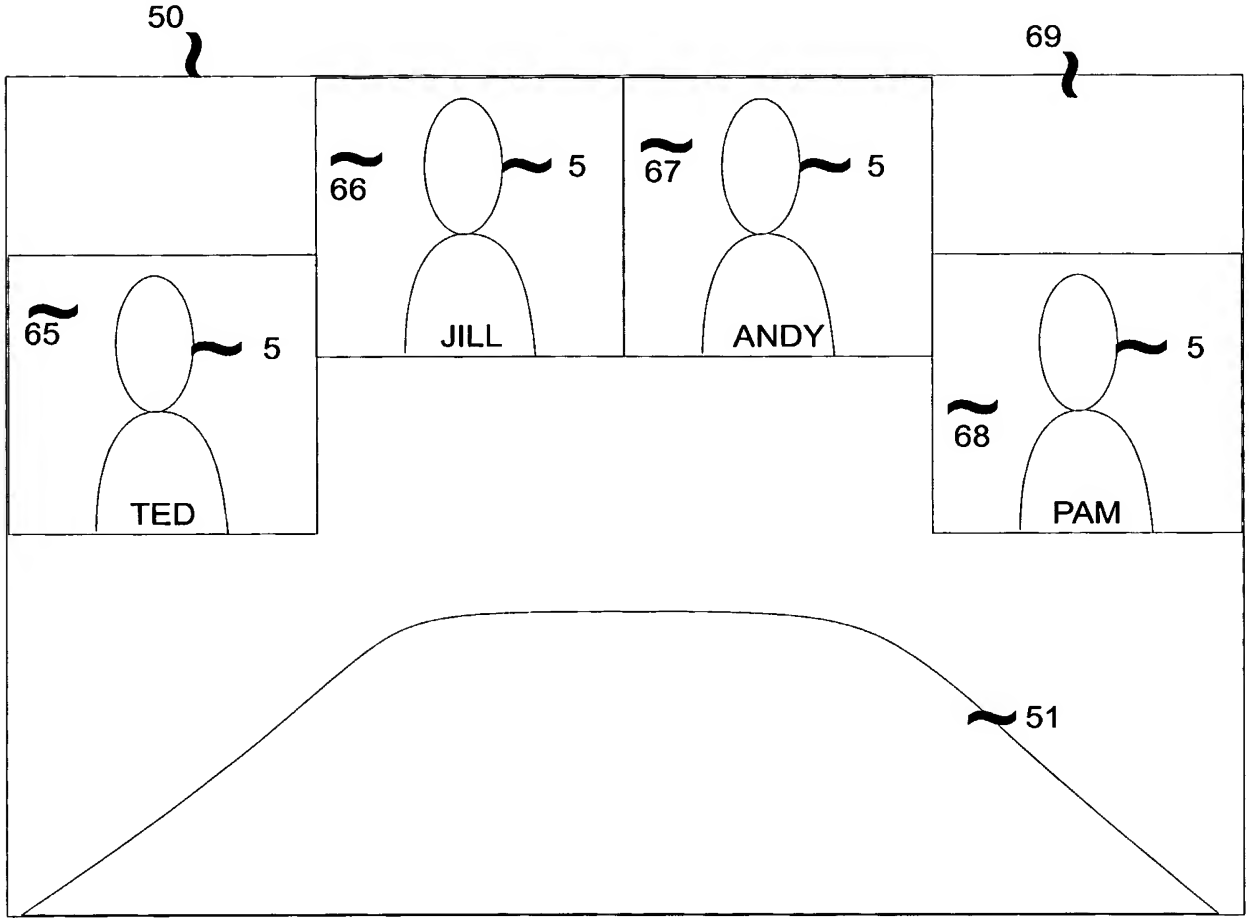
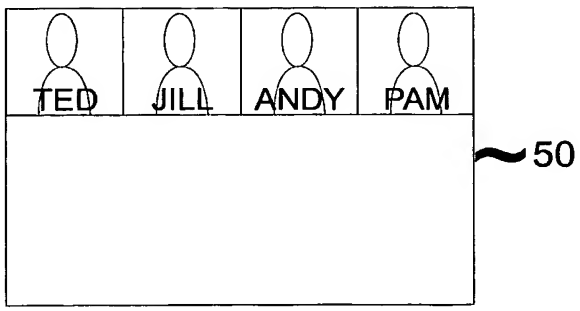


Fig 17a



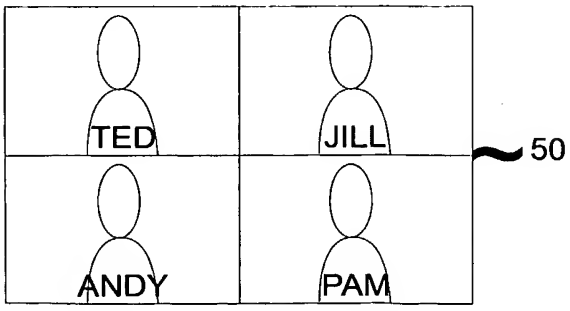
LAYOUT 1

Fig 17b



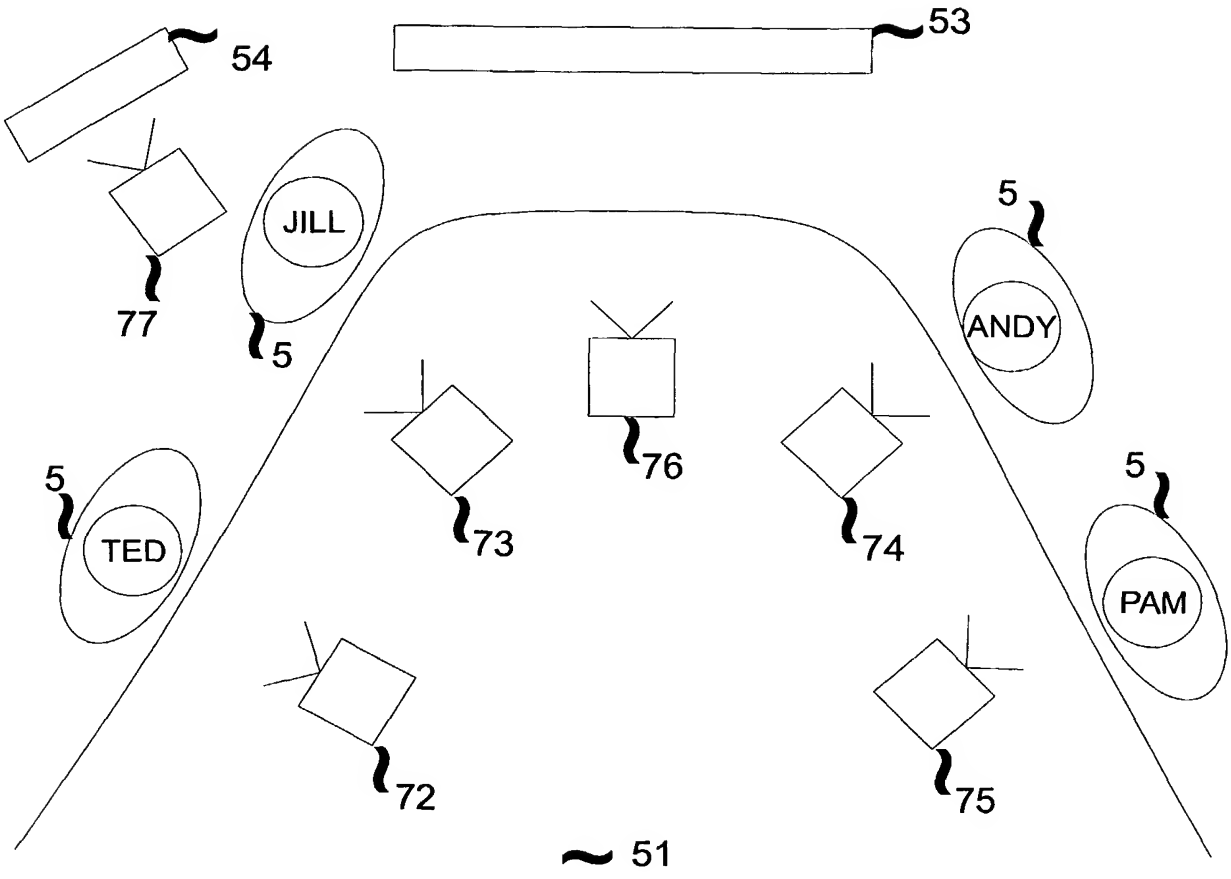
LAYOUT 2

Fig 17c



LAYOUT 3

Fig 18



MEETING ROOM

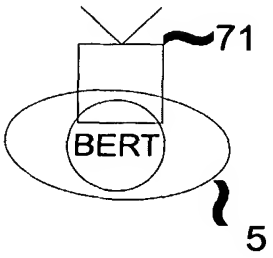


Fig 19

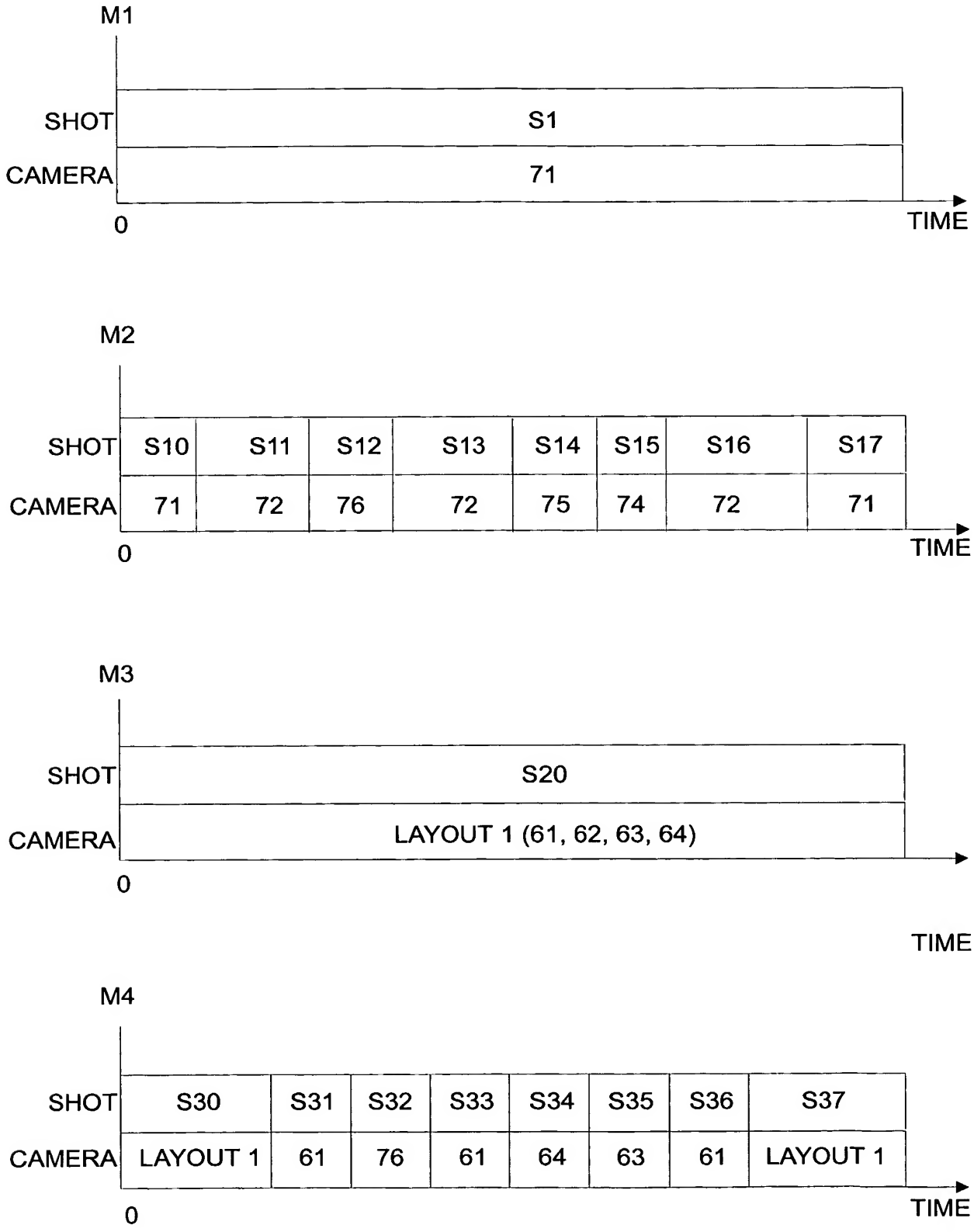

M1
SHOT
S1
CAMERA
71
0
TIME
M2
SHOT
S10
S11
S12
S13
S14
S15
S16
S17
CAMERA
71
72
76
72
75
74
72
71
0
TIME
M3
SHOT
S20
CAMERA
LAYOUT 1 (61, 62, 63, 64)
0
TIME
M4
SHOT
S30
S31
S32
S33
S34
S35
S36
S37
CAMERA
LAYOUT 1
61
76
61
64
63
61
LAYOUT 1
0
TIME

Fig 20

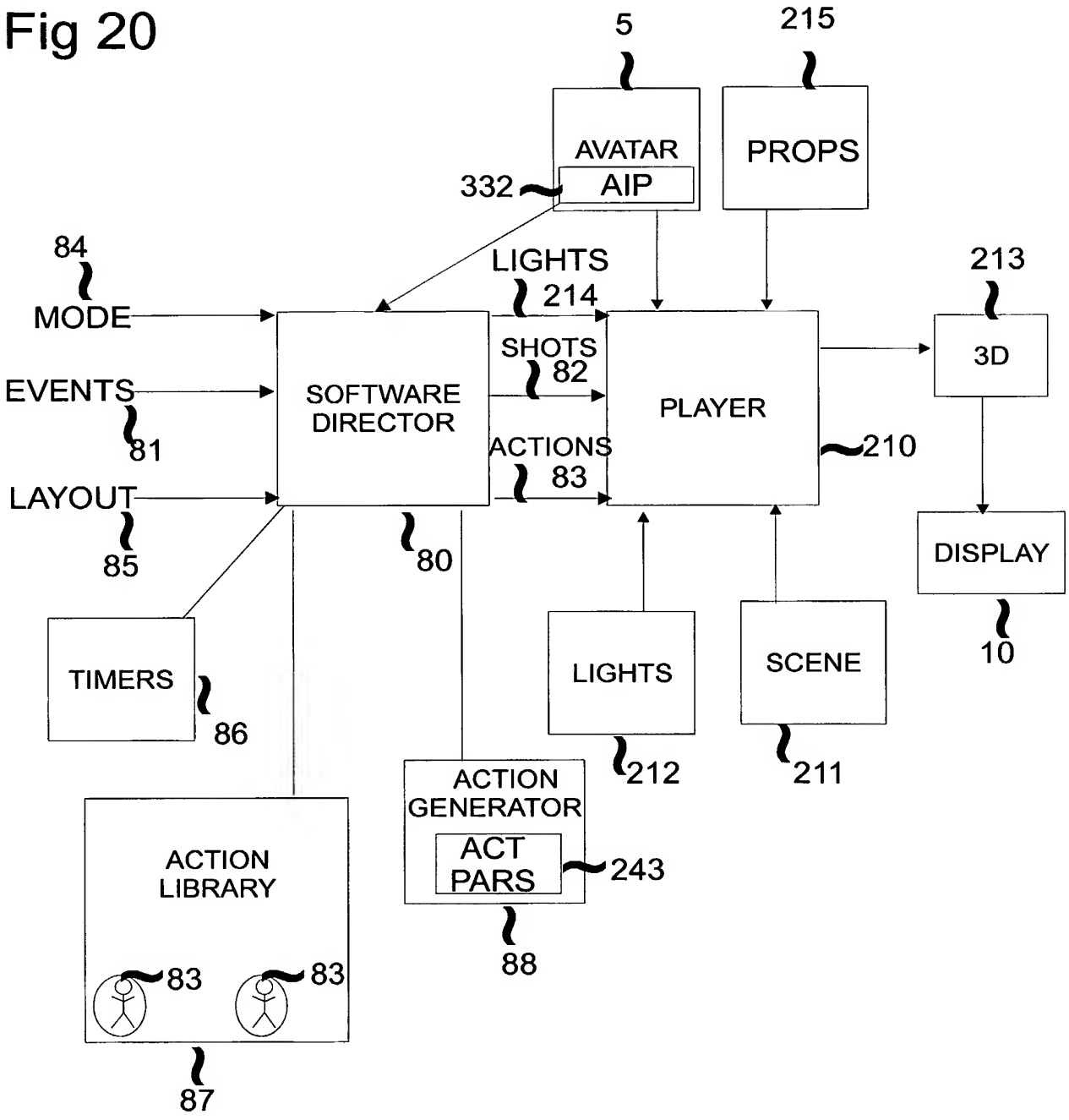
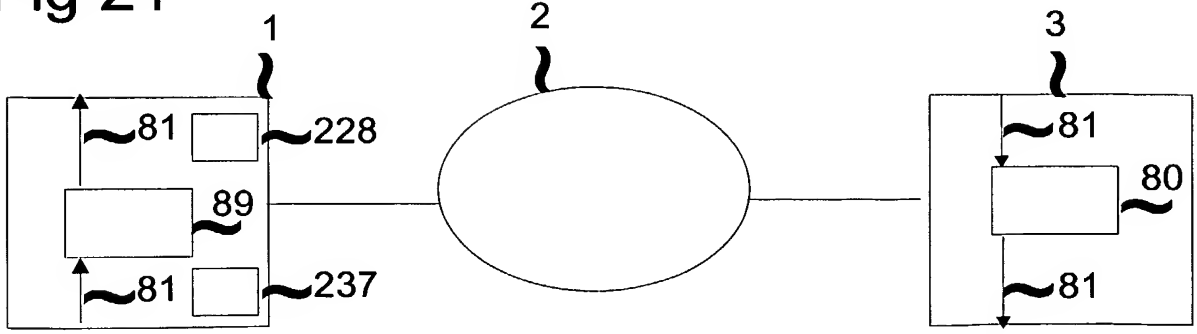

5
215
AVATAR
AIP
PROPS
332
84
MODE
LIGHTS
214
213
EVENTS
81
SOFTWARE DIRECTOR
SHOTS
82
PLAYER
3D
ACTIONS
83
210
LAYOUT
85
80
DISPLAY
10
TIMERS
86
LIGHTS
212
SCENE
211
ACTION GENERATOR
ACT PARS
243
88
ACTION LIBRARY
83
83
87

Fig 21


1
2
3
81
228
89
81
237
81
80
81

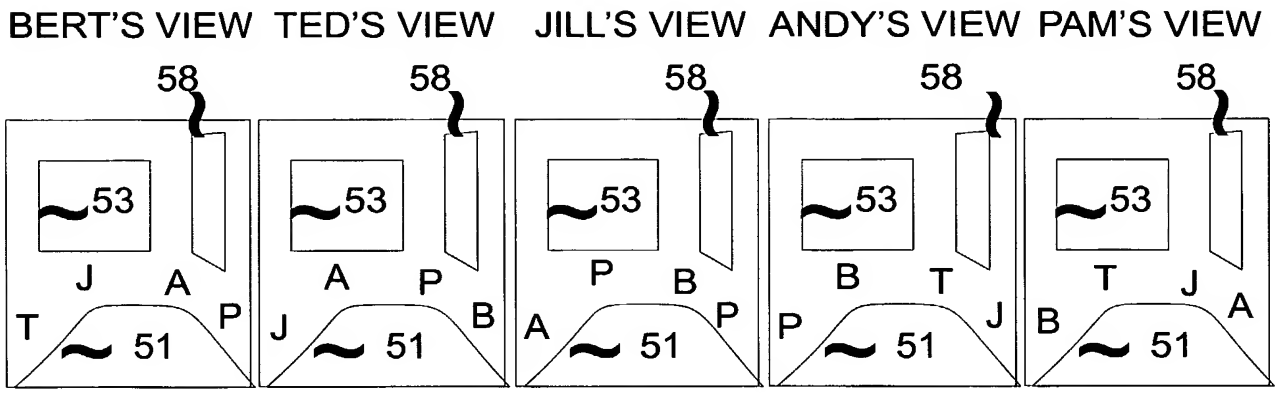


Fig 22a Fig 22b Fig 22c Fig 22d Fig 22e

Fig 23

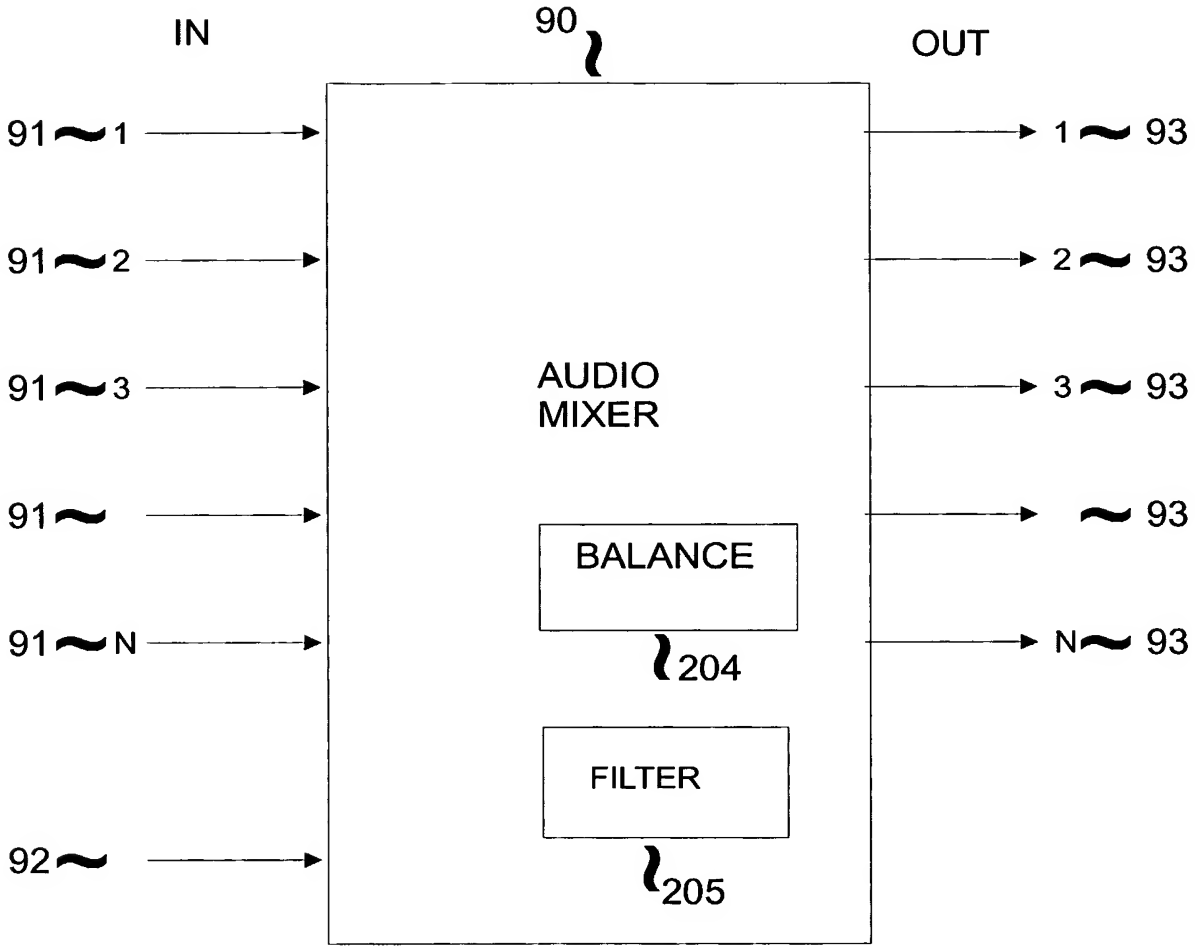

IN
90
OUT
91 1
91 2
91 3
91
91 N
92
AUDIO MIXER
BALANCE
204
FILTER
205
1 93
2 93
3 93
93
N 93

Fig 24

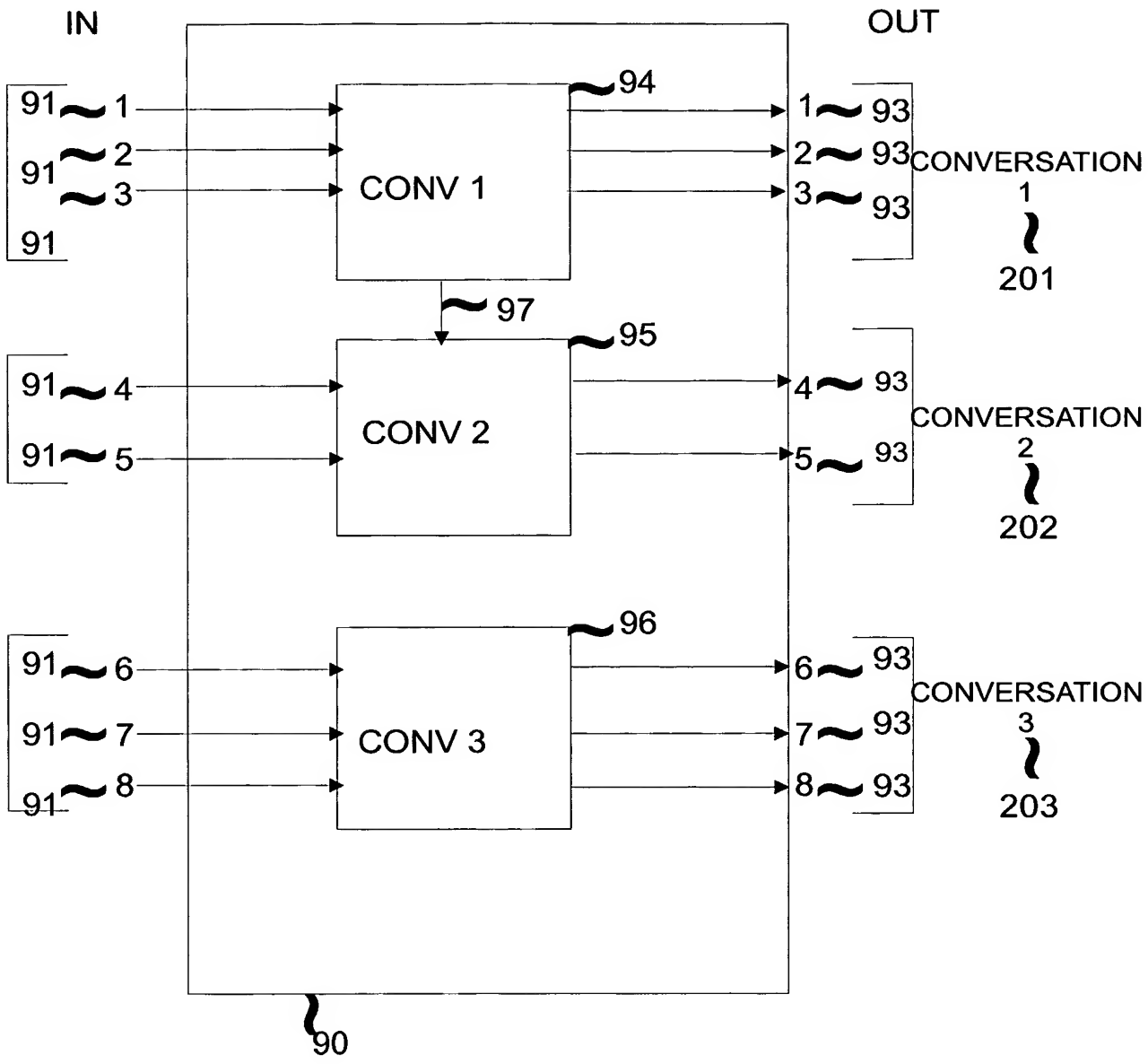

IN
OUT
91
1
2
3
CONV 1
94
93
CONVERSATION 1
201
97
95
4
5
CONV 2
CONVERSATION 2
202
96
6
7
8
CONV 3
CONVERSATION 3
203
90

Fig 25

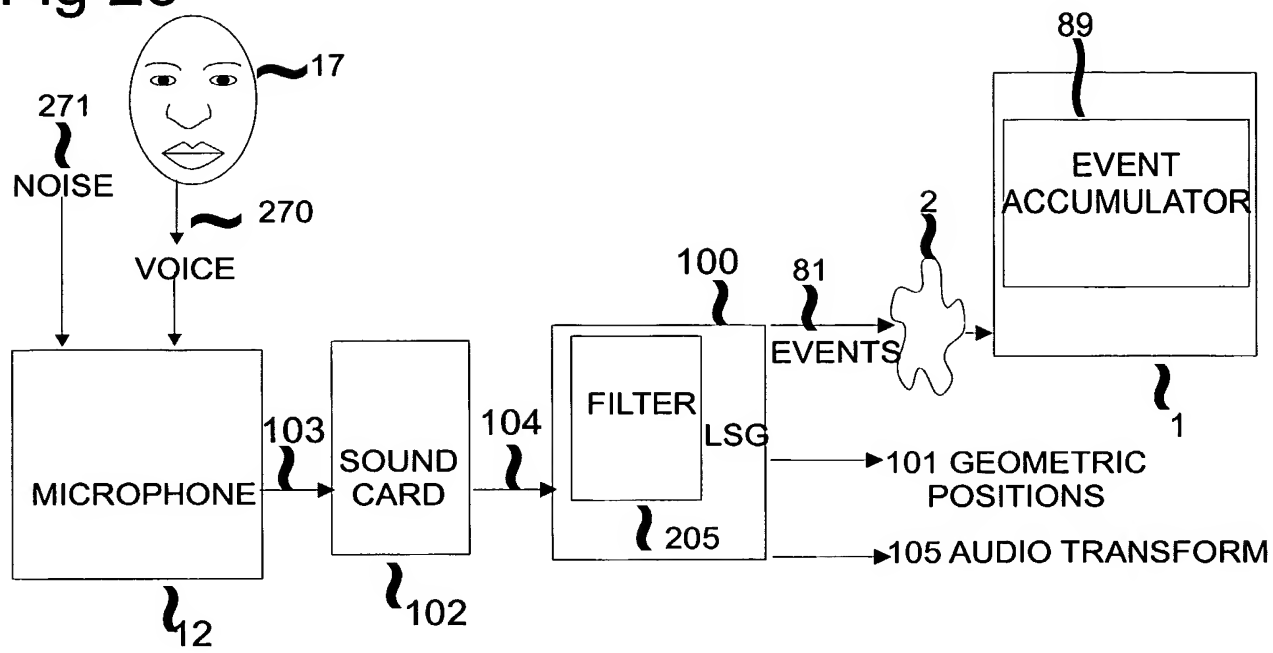

17
271
NOISE
270
VOICE
89
EVENT ACCUMULATOR
2
100
81
EVENTS
MICROPHONE
103
SOUND CARD
104
FILTER
LSG
205
1
101 GEOMETRIC POSITIONS
105 AUDIO TRANSFORM
12
102

Fig 26

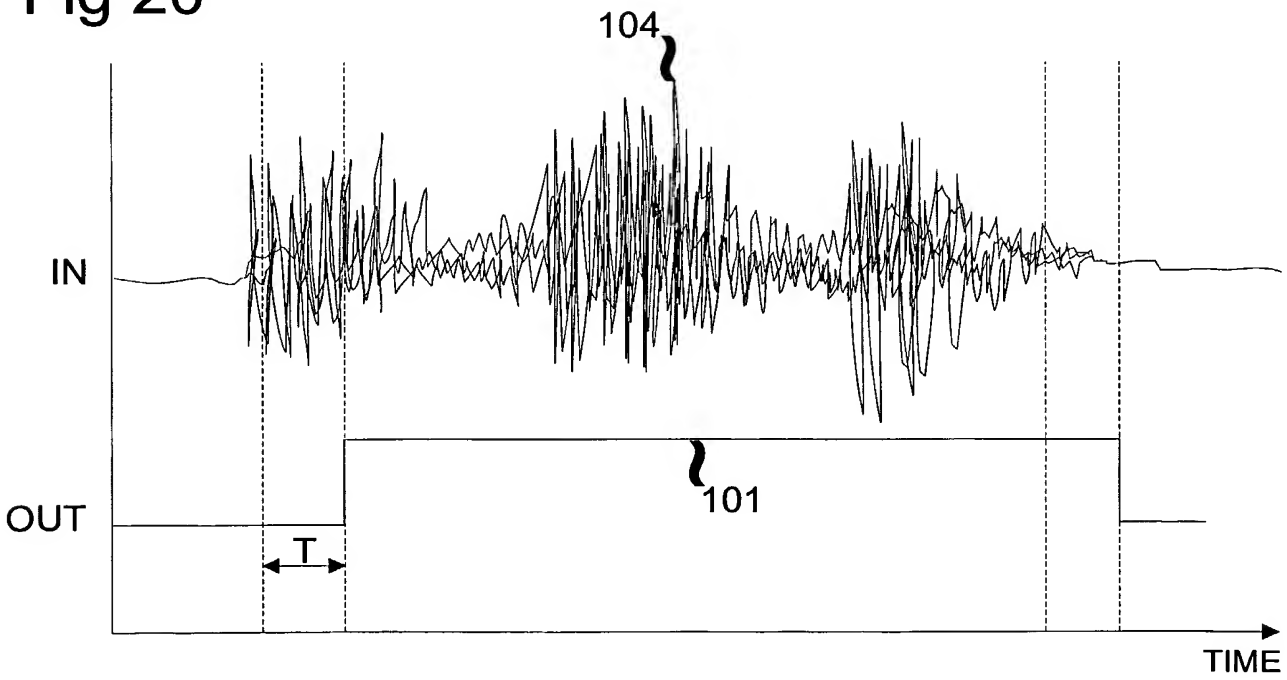

104
IN
OUT
101
T
TIME

Fig 27a

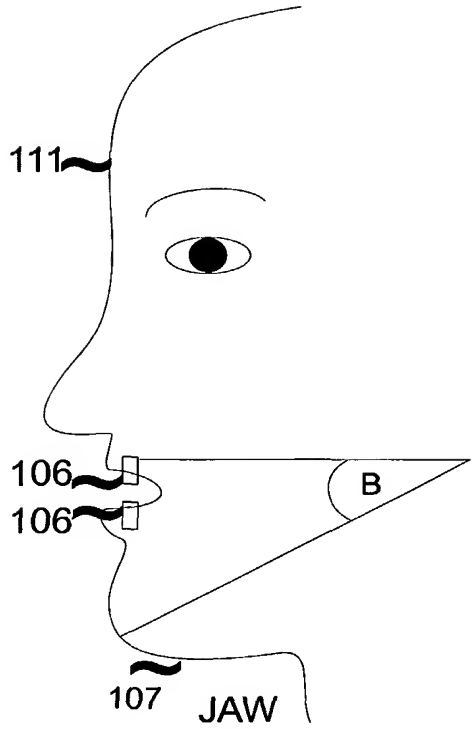


Fig 27b

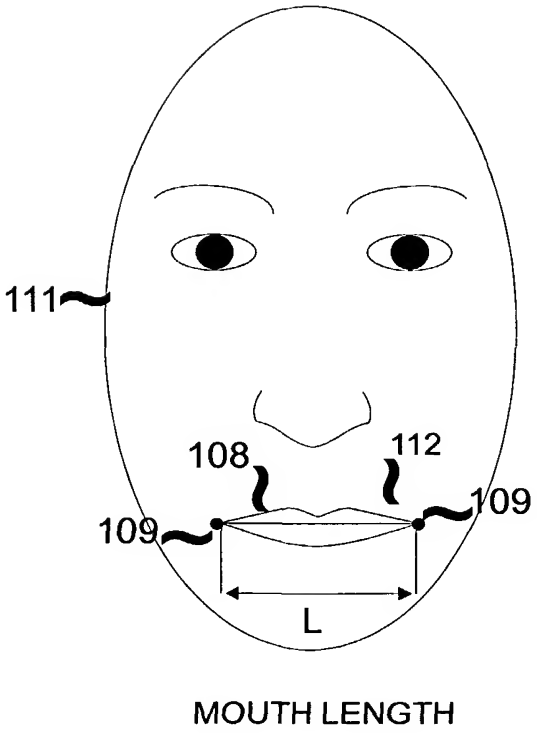


Fig 27c

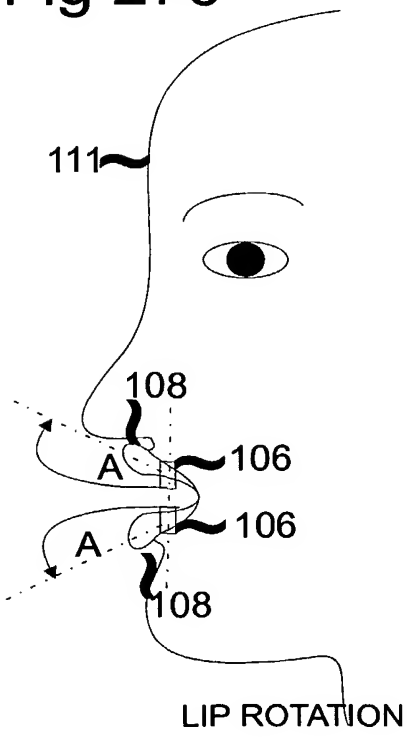


Fig 27d

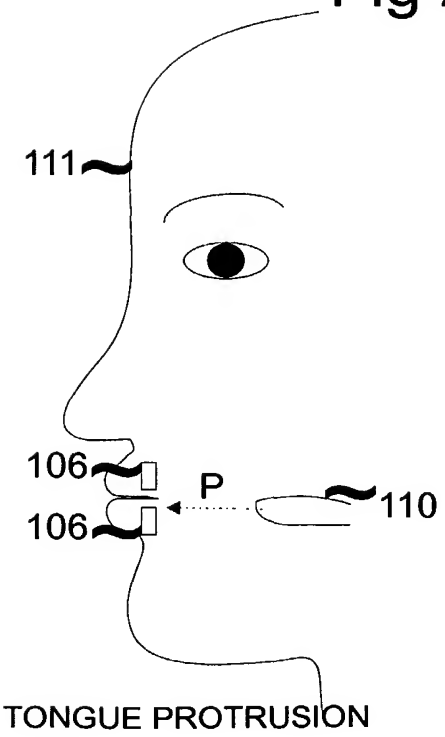


Fig 28

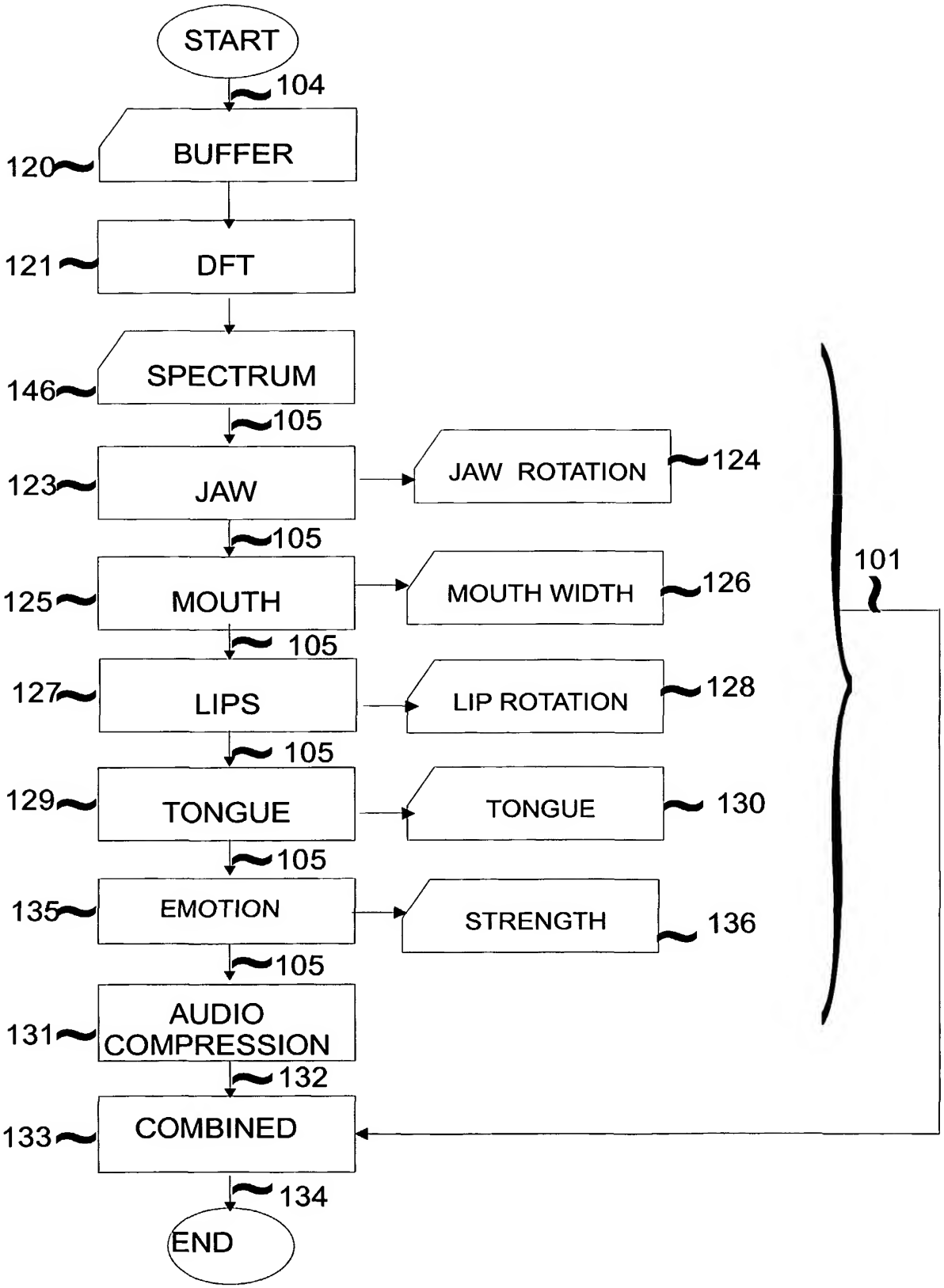

START
104
BUFFER
120
DFT
121
SPECTRUM
146
105
JAW
123
JAW ROTATION
124
105
MOUTH
125
MOUTH WIDTH
126
105
LIPS
127
LIP ROTATION
128
105
TONGUE
129
TONGUE
130
105
EMOTION
135
STRENGTH
136
105
AUDIO COMPRESSION
131
132
COMBINED
133
134
END
101

Fig 29

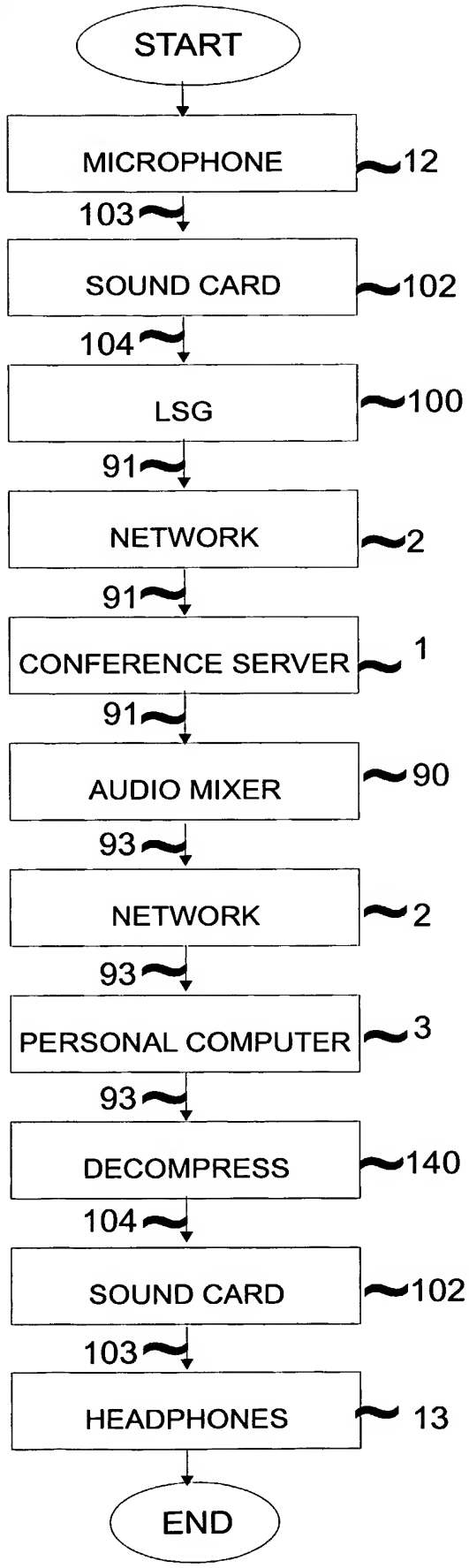

START
MICROPHONE
12
103
SOUND CARD
102
104
LSG
100
91
NETWORK
2
91
CONFERENCE SERVER
1
91
AUDIO MIXER
90
93
NETWORK
2
93
PERSONAL COMPUTER
3
93
DECOMPRESS
140
104
SOUND CARD
102
103
HEADPHONES
13
END

Fig 30a

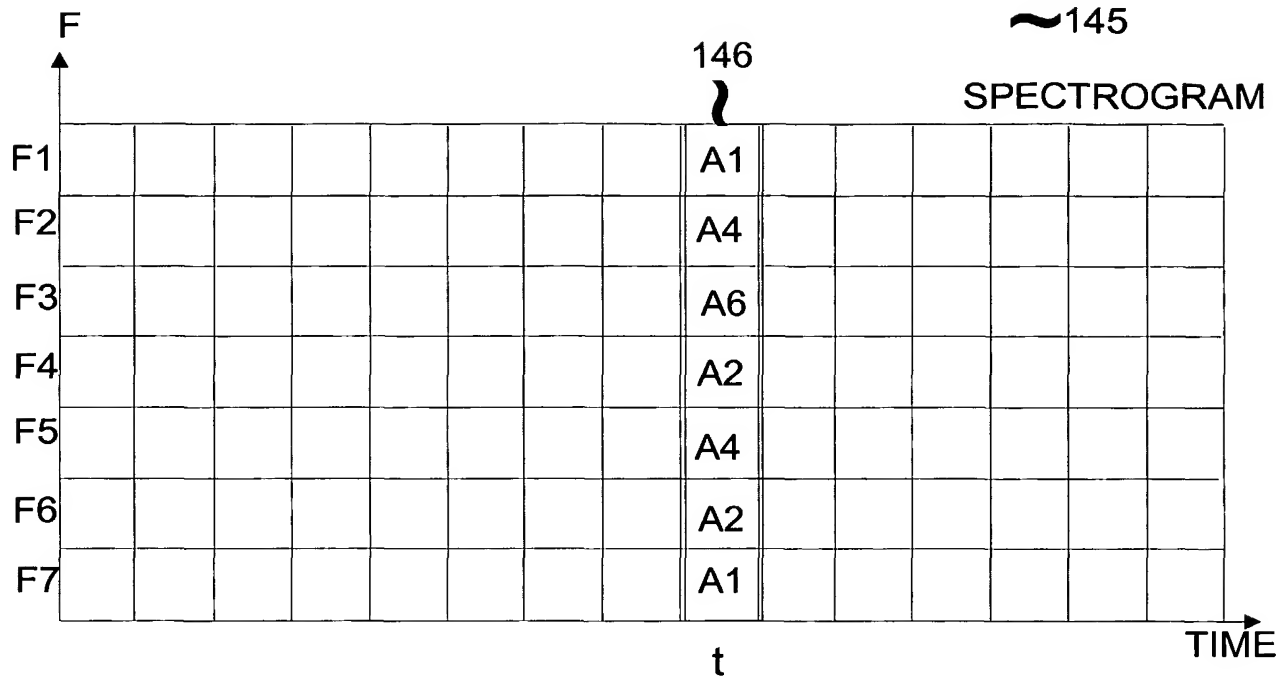


Fig 30b

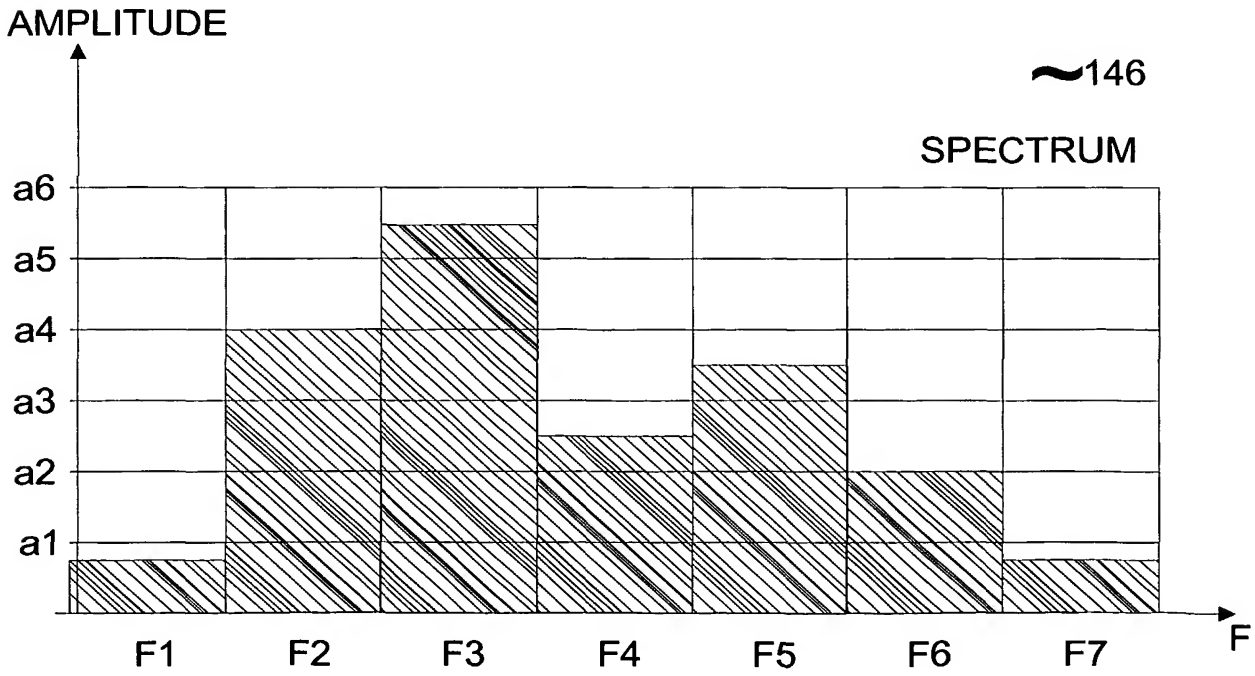


Fig 31

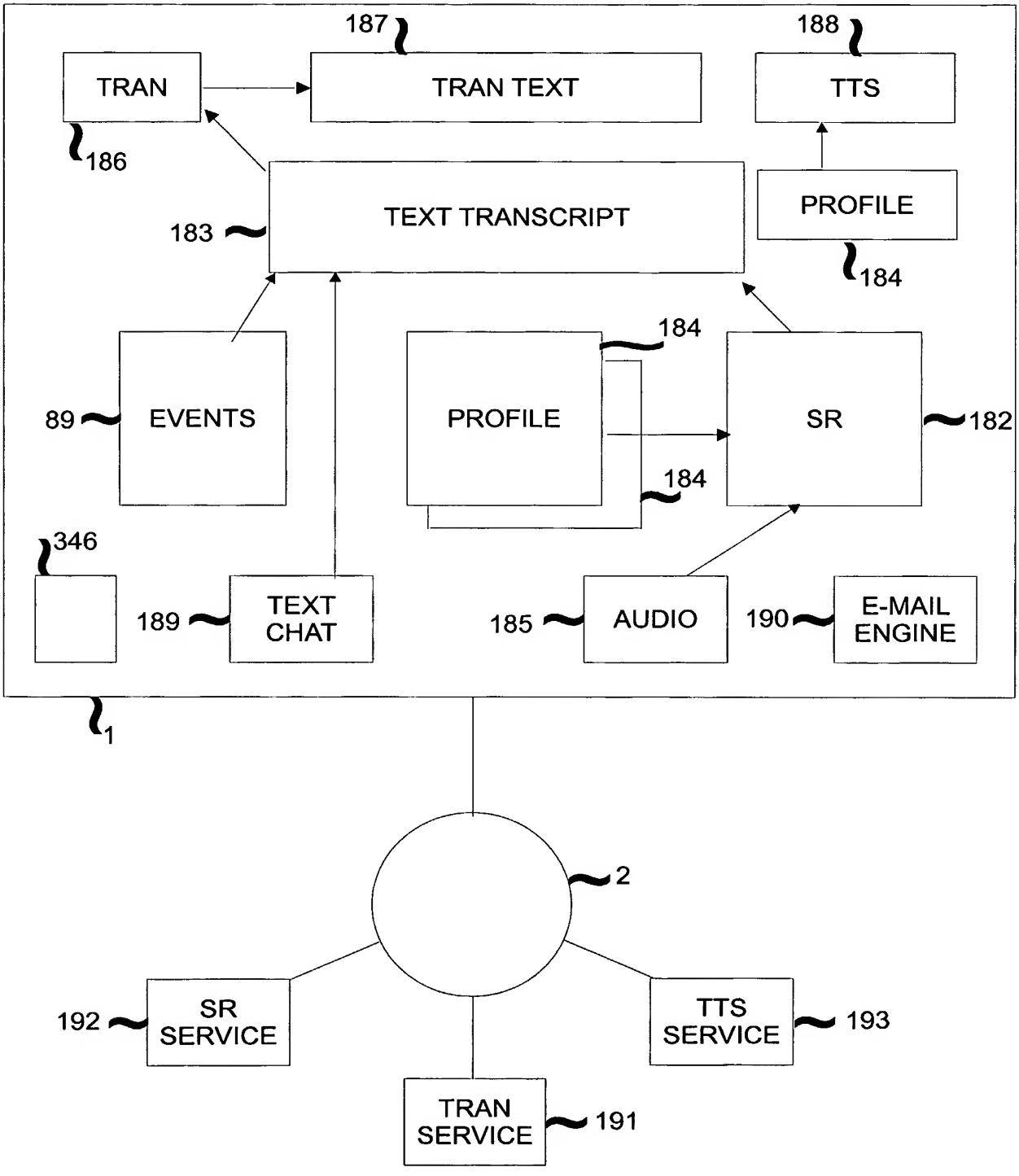

187
188
TRAN
TRAN TEXT
TTS
186
TEXT TRANSCRIPT
PROFILE
183
184
184
89
EVENTS
PROFILE
SR
182
184
346
189
TEXT CHAT
185
AUDIO
190
E-MAIL ENGINE
1
2
192
SR SERVICE
TTS SERVICE
193
TRAN SERVICE
191

Fig 32

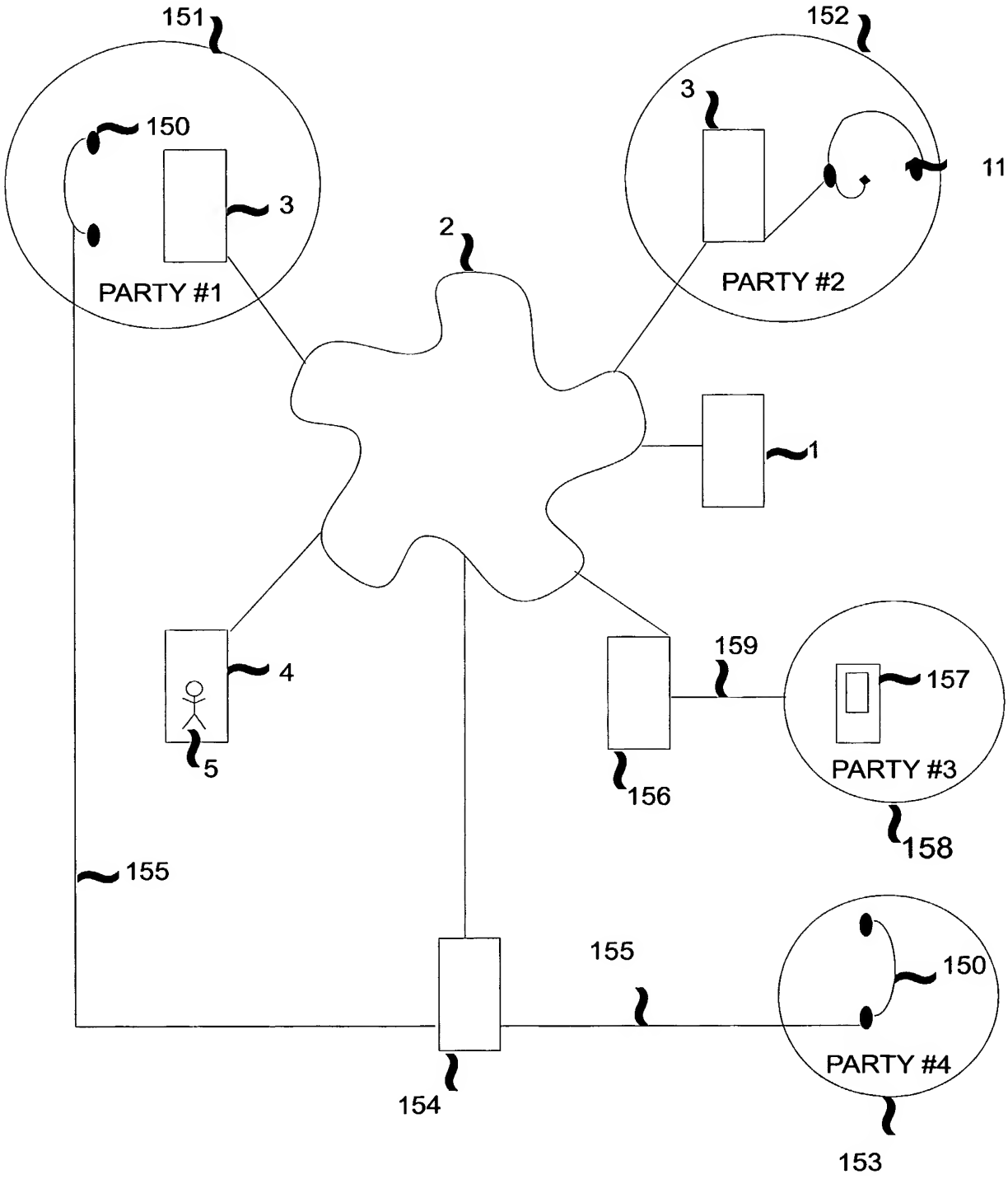

151
152
150
3
3
11
2
PARTY #1
PARTY #2
1
4
5
159
157
156
PARTY #3
158
155
155
150
154
PARTY #4
153

Fig 33

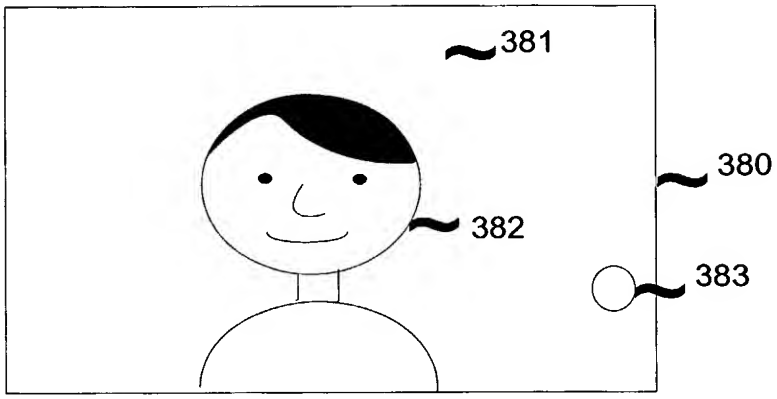


Fig 34

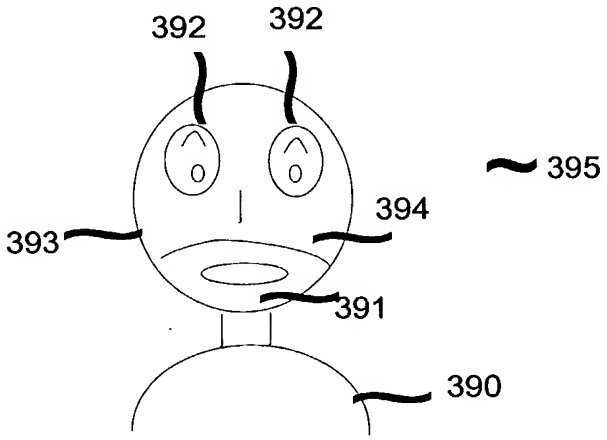


Fig 35

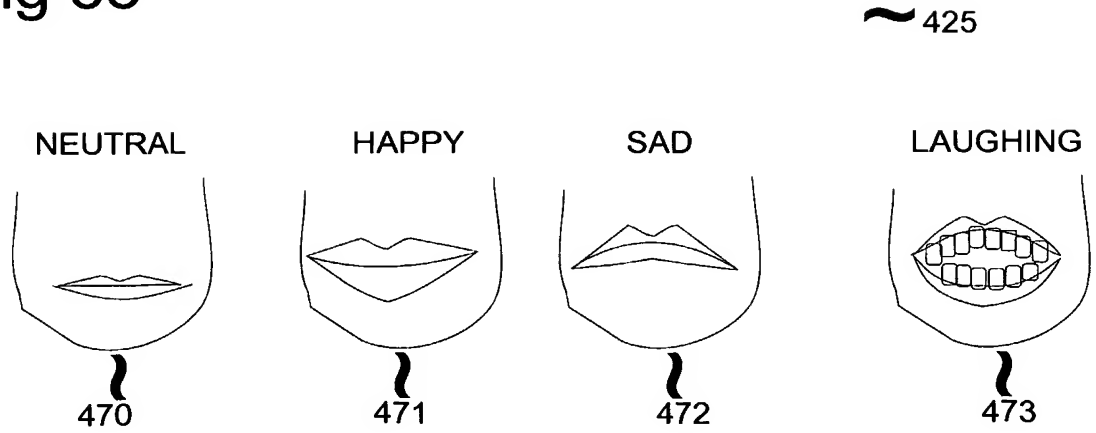


Fig 36

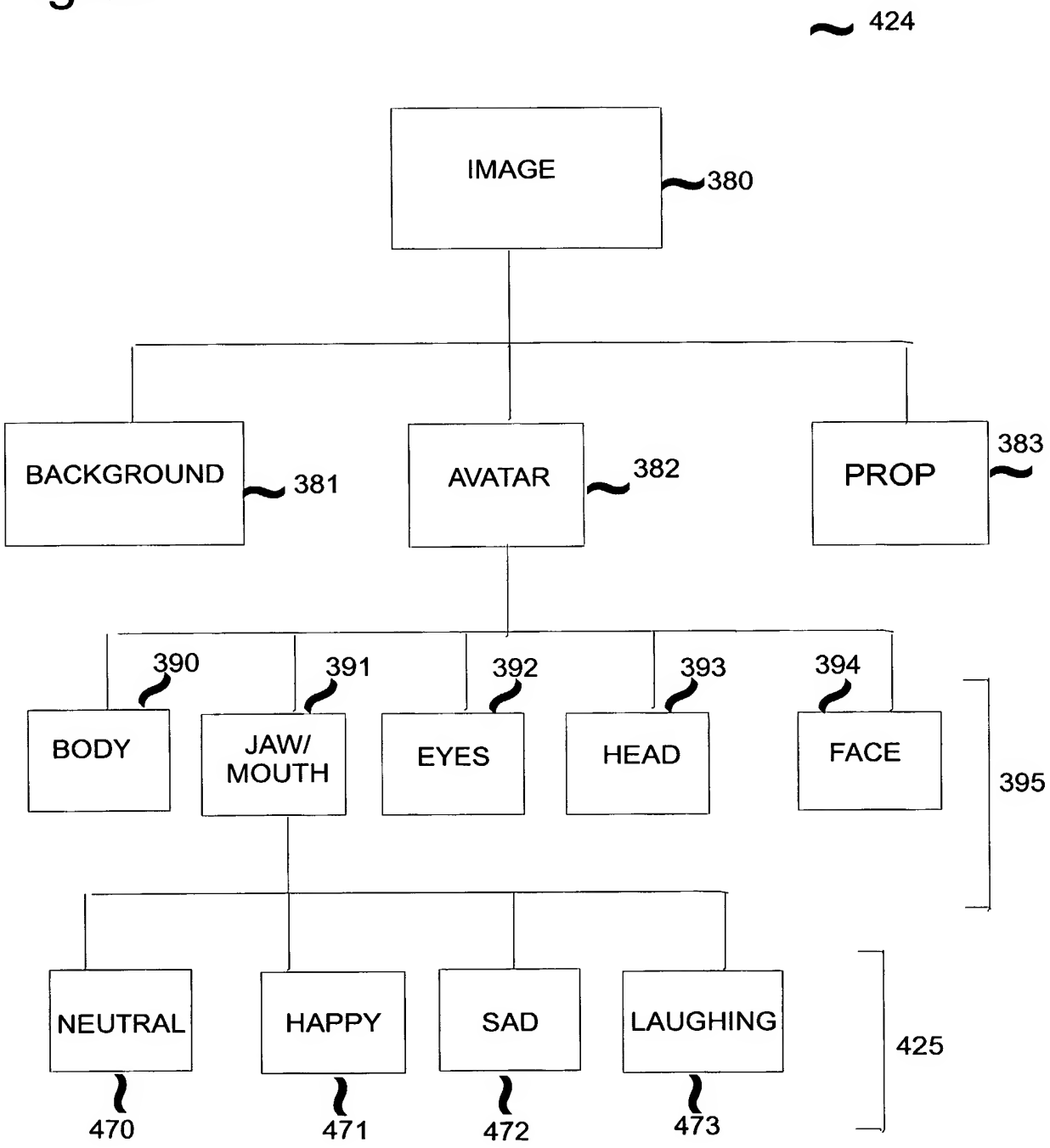

424
IMAGE
380
BACKGROUND
381
AVATAR
382
PROP
383
390
391
392
393
394
BODY
JAW/ MOUTH
EYES
HEAD
FACE
395
NEUTRAL
HAPPY
SAD
LAUGHING
425
470
471
472
473

Fig 37

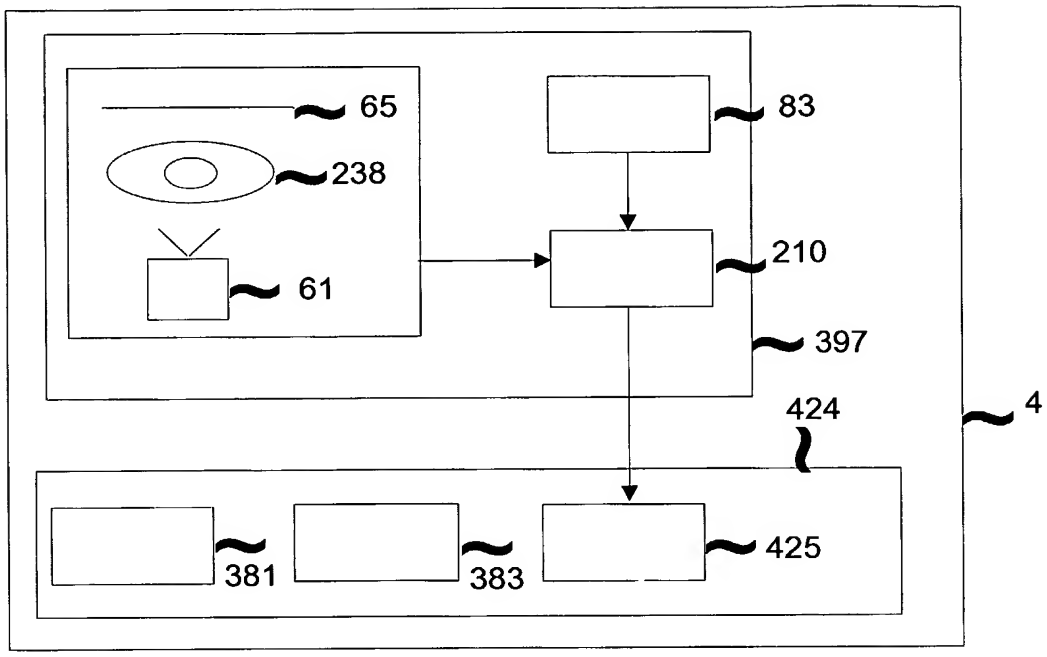

65
238
61
83
210
397
424
4
381
383
425

Fig 38

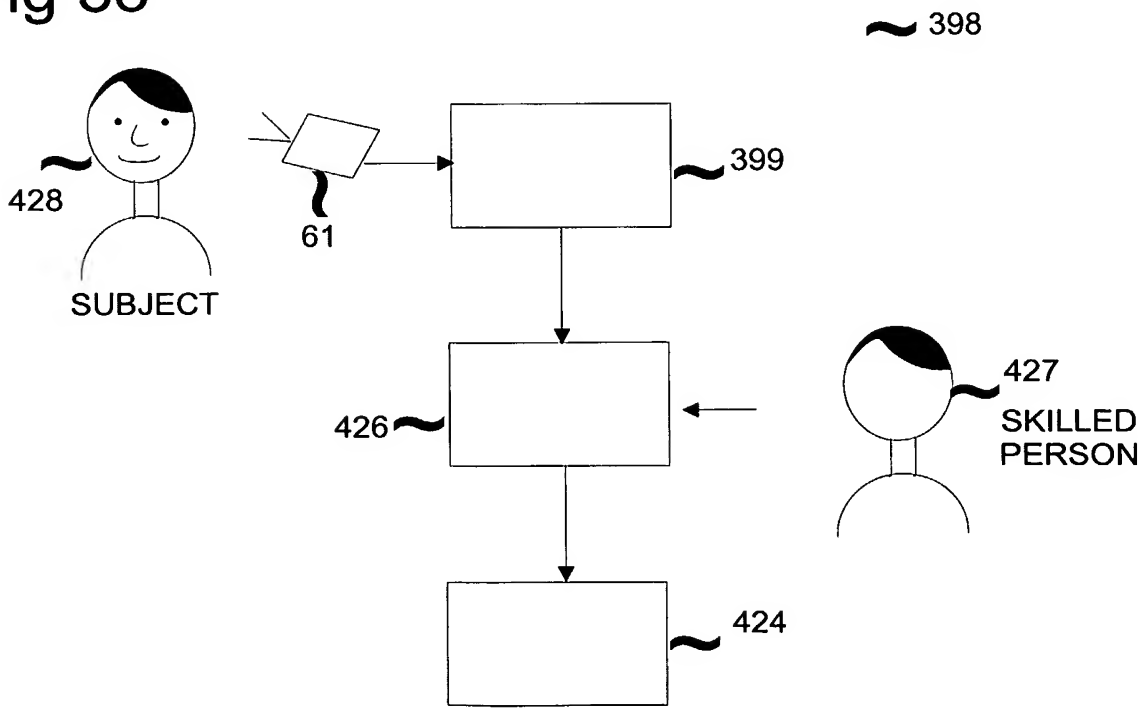

398
428
SUBJECT
61
399
426
427
SKILLED PERSON
424

Fig 39

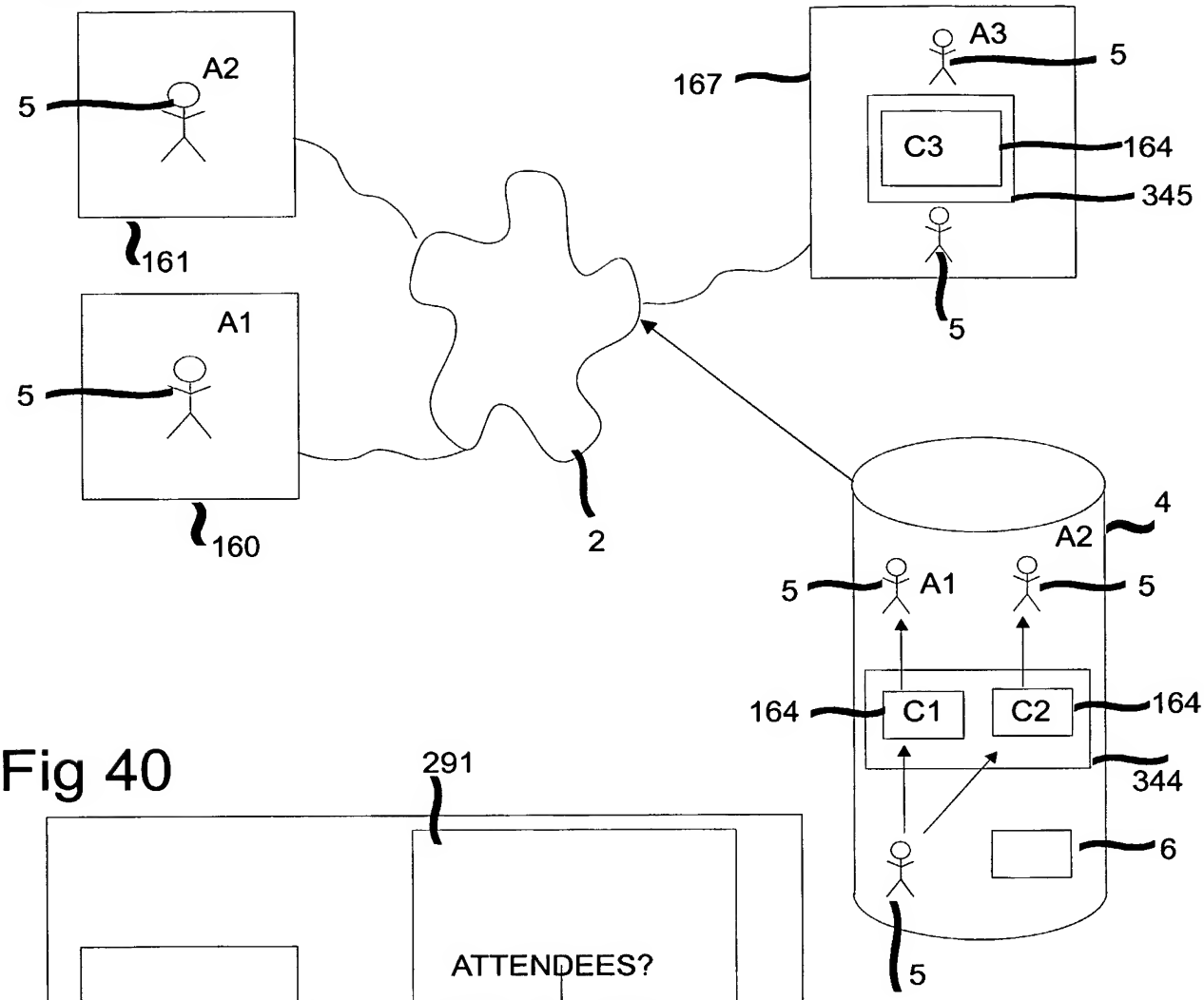

A2
5
161
A1
5
160
2
A3
167
5
C3
164
345
5
4
A2
5
A1
5
164
C1
C2
164
344
6
5

Fig 40

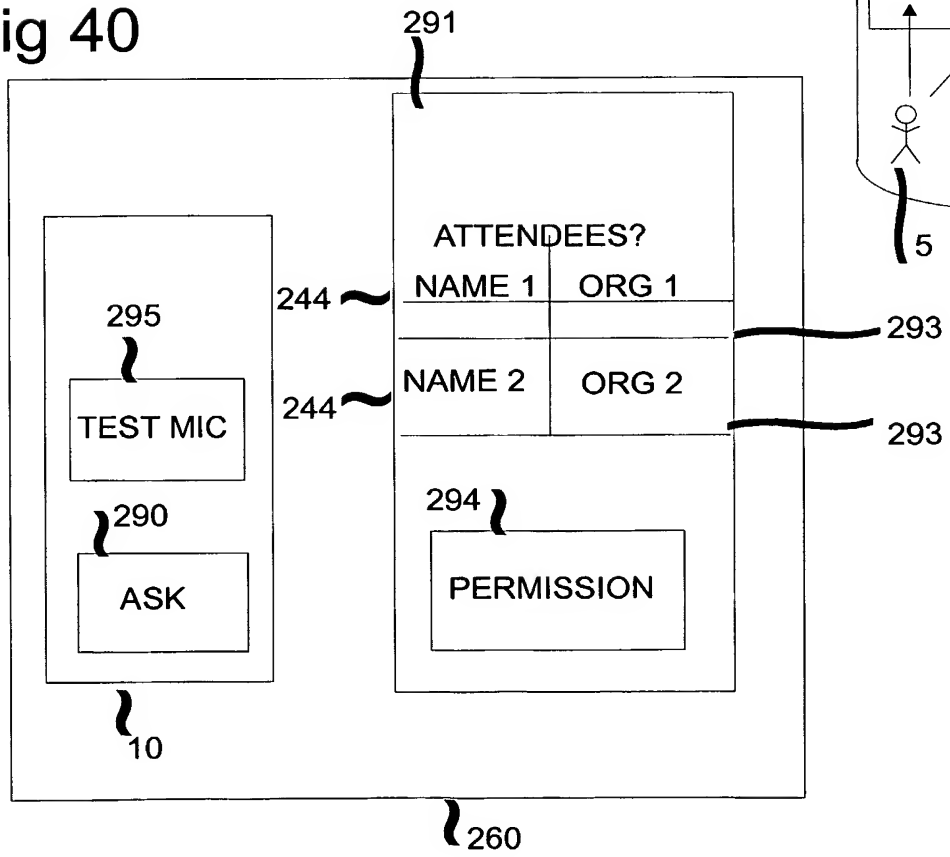

291
ATTENDEES?
NAME 1
ORG 1
244
293
NAME 2
ORG 2
244
293
294
PERMISSION
295
TEST MIC
290
ASK
10
260

Fig 41

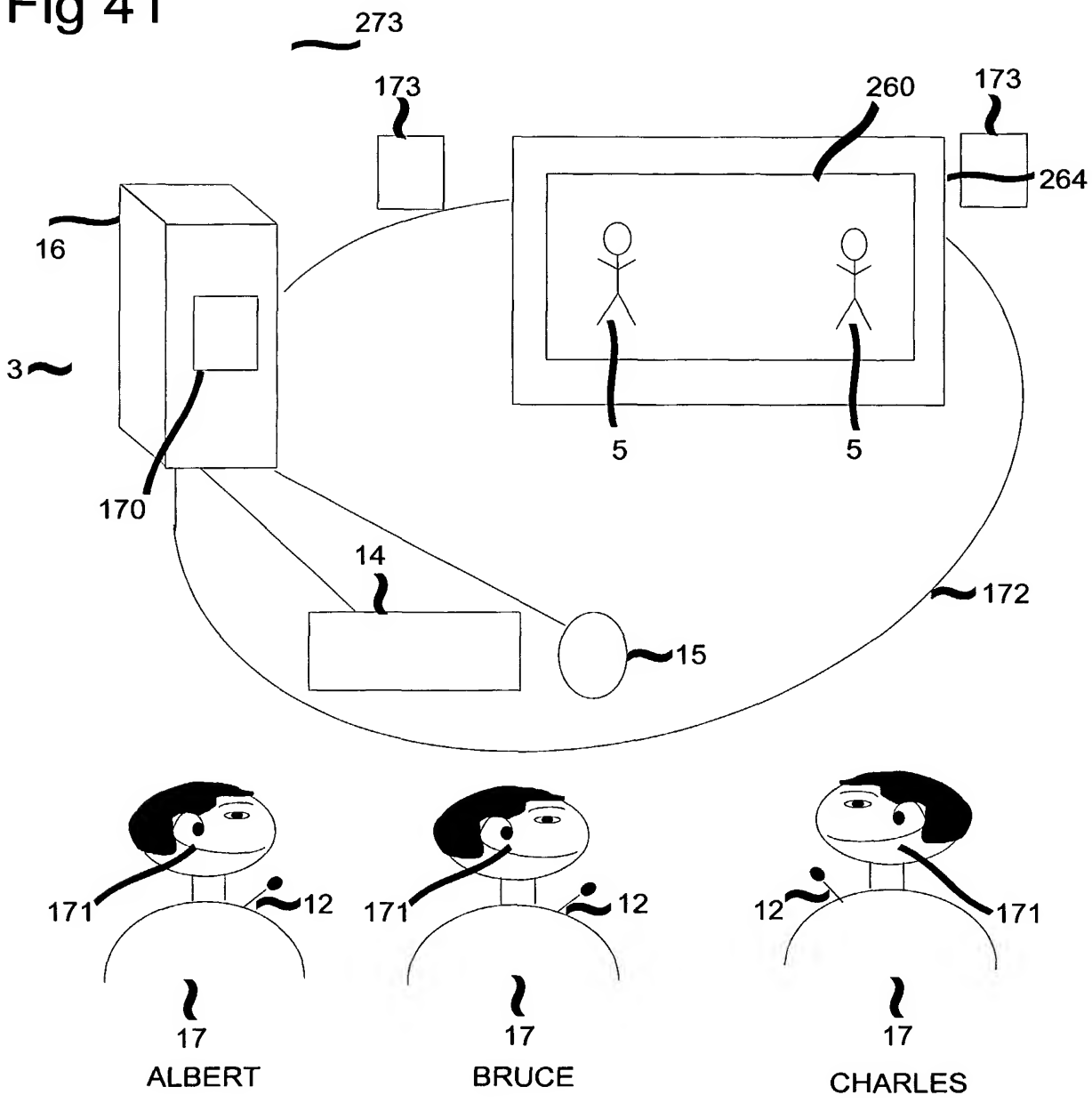

273
173
260
173
264
16
3
5
5
170
14
172
15
171
12
17
ALBERT
171
12
17
BRUCE
12
171
17
CHARLES

Fig 41a

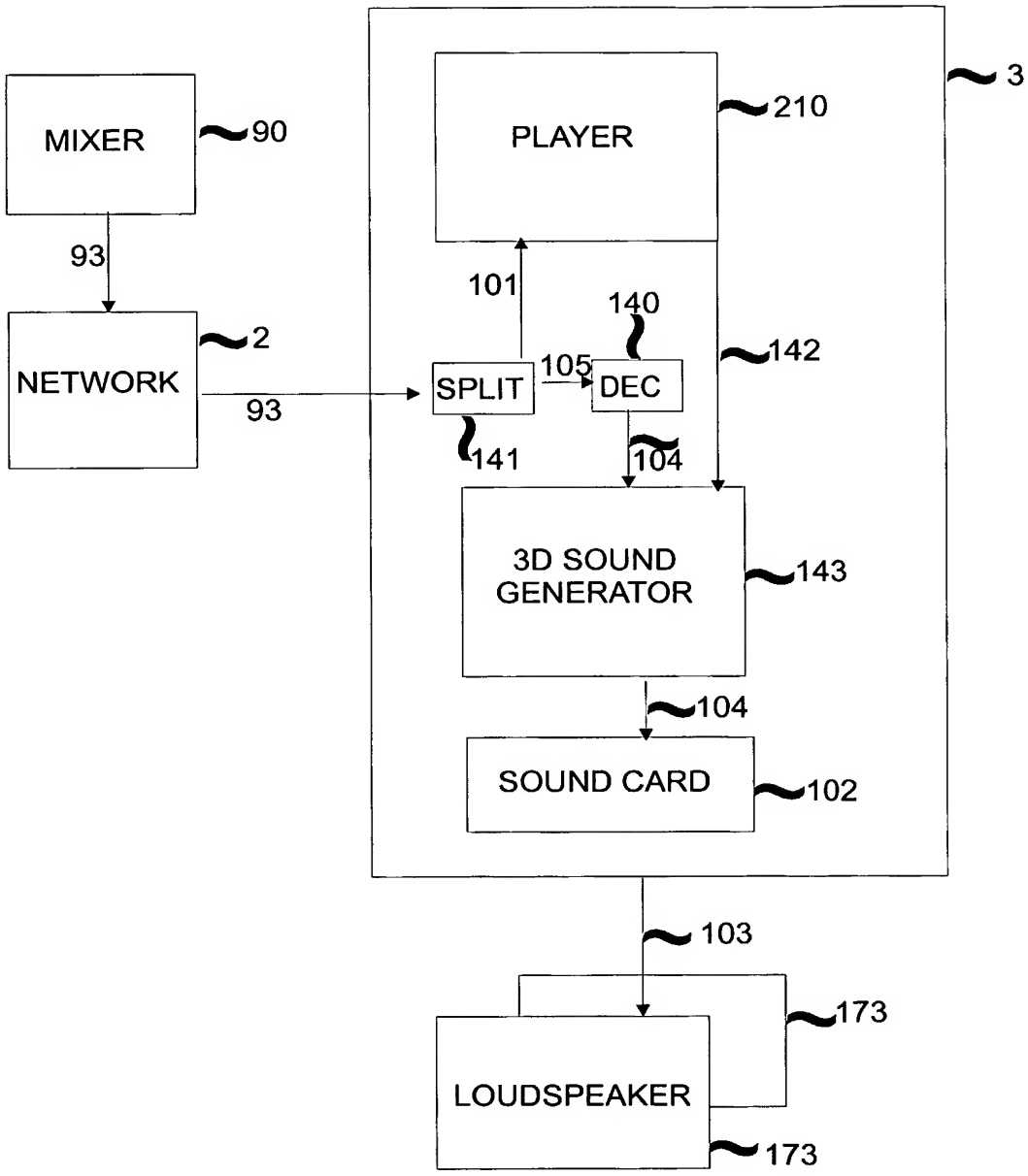

MIXER
90
93
NETWORK
2
93
3
PLAYER
210
101
140
SPLIT
105
DEC
142
141
104
3D SOUND GENERATOR
143
104
SOUND CARD
102
103
173
LOUDSPEAKER
173

Fig 42

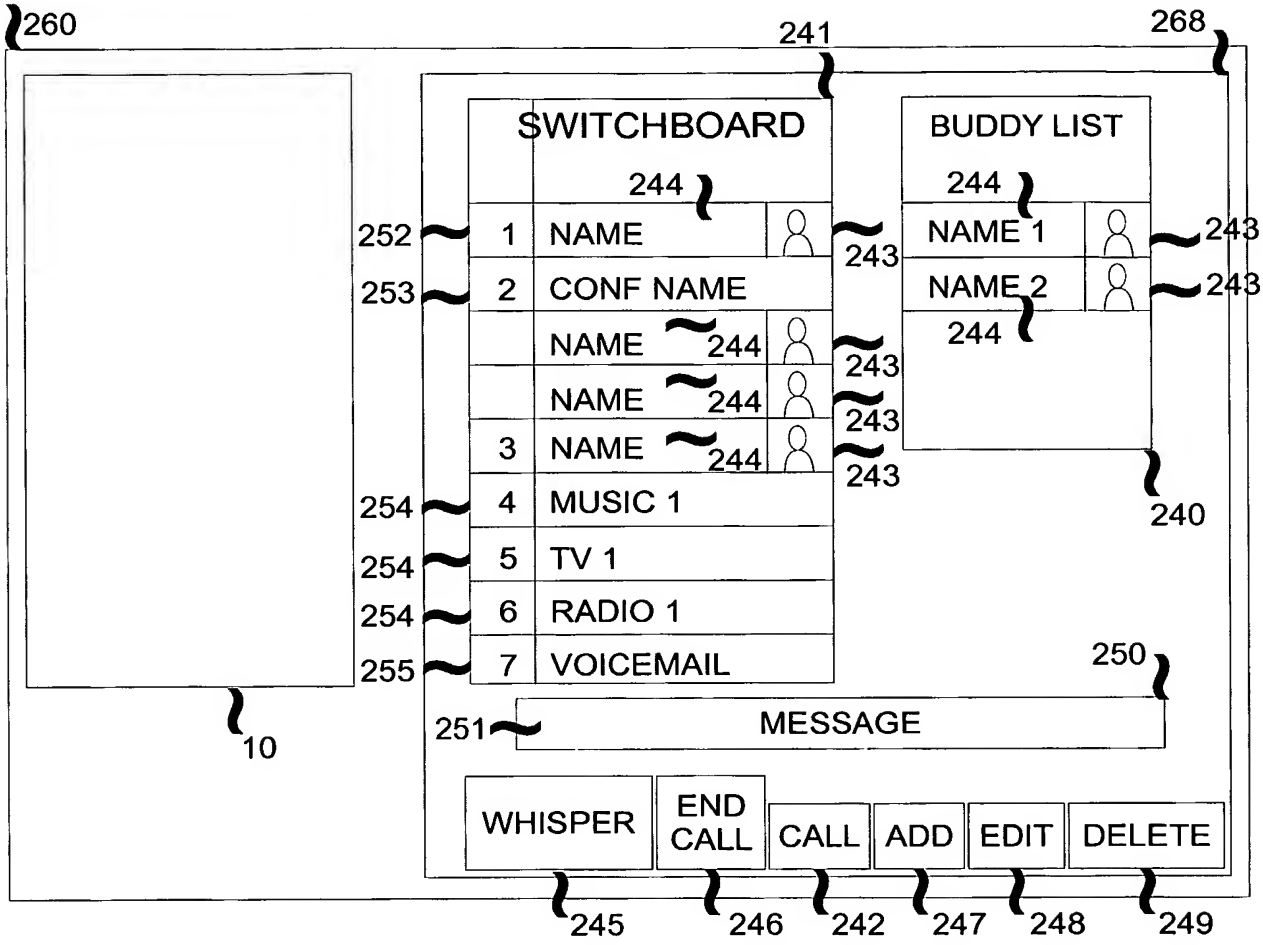

260
241
268
SWITCHBOARD
BUDDY LIST
244
244
252
1 NAME
243
NAME 1
243
253
2 CONF NAME
NAME 2
243
244
NAME 244
243
NAME 244
243
3 NAME 244
243
240
254
4 MUSIC 1
254
5 TV 1
254
6 RADIO 1
255
7 VOICEMAIL
250
251
MESSAGE
10
WHISPER
END CALL
CALL
ADD
EDIT
DELETE
245
246
242
247
248
249

Fig 43

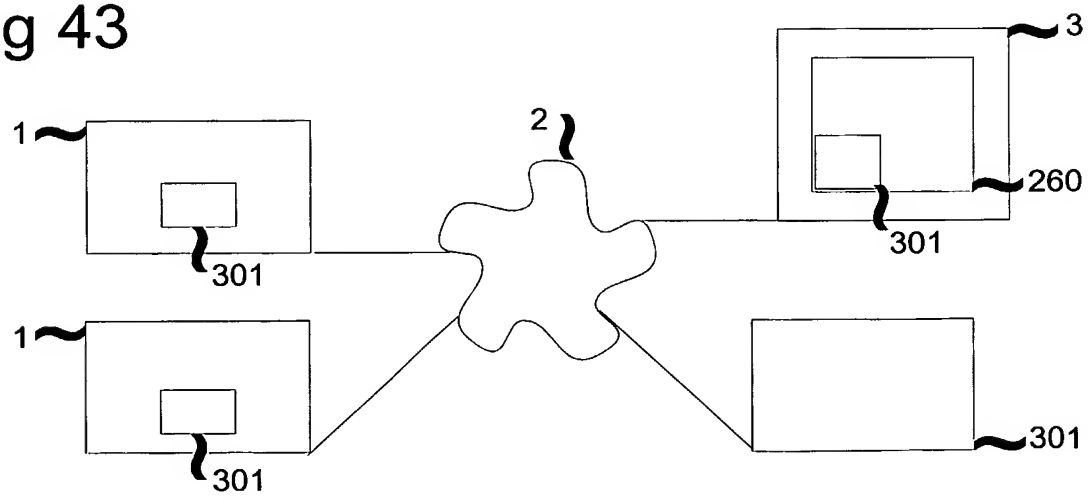

1
2
3
260
301
1
301
301

Fig 44

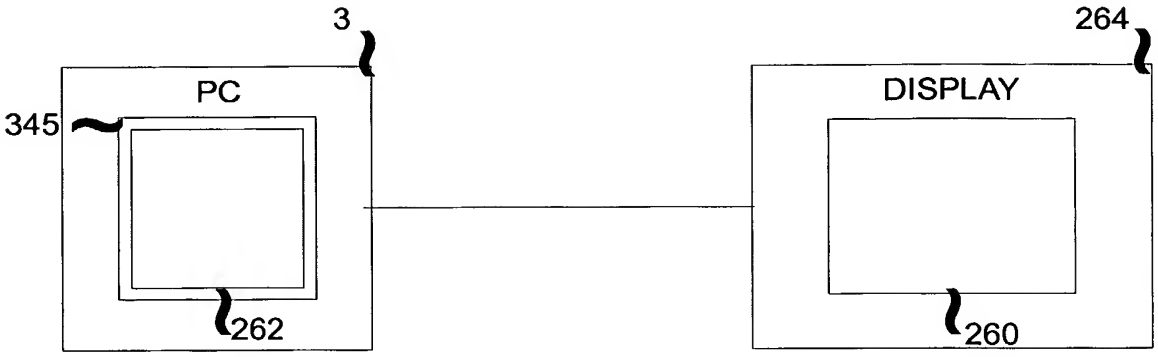

3
PC
345
262
264
DISPLAY
260

Fig 45

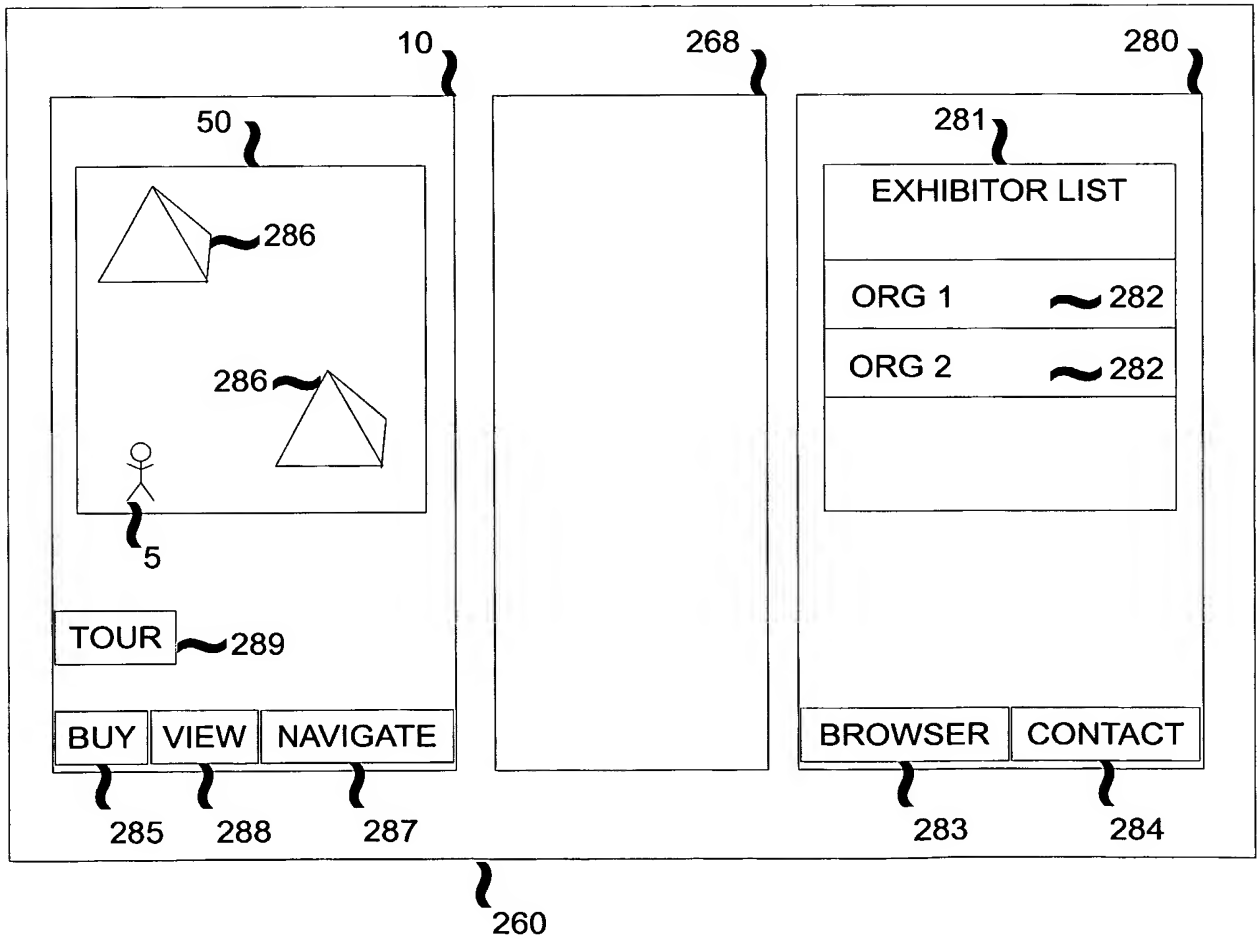

10
50
286
286
5
TOUR
289
BUY
VIEW
NAVIGATE
285
288
287
268
280
281
EXHIBITOR LIST
ORG 1
282
ORG 2
282
BROWSER
CONTACT
283
284
260

Fig 46

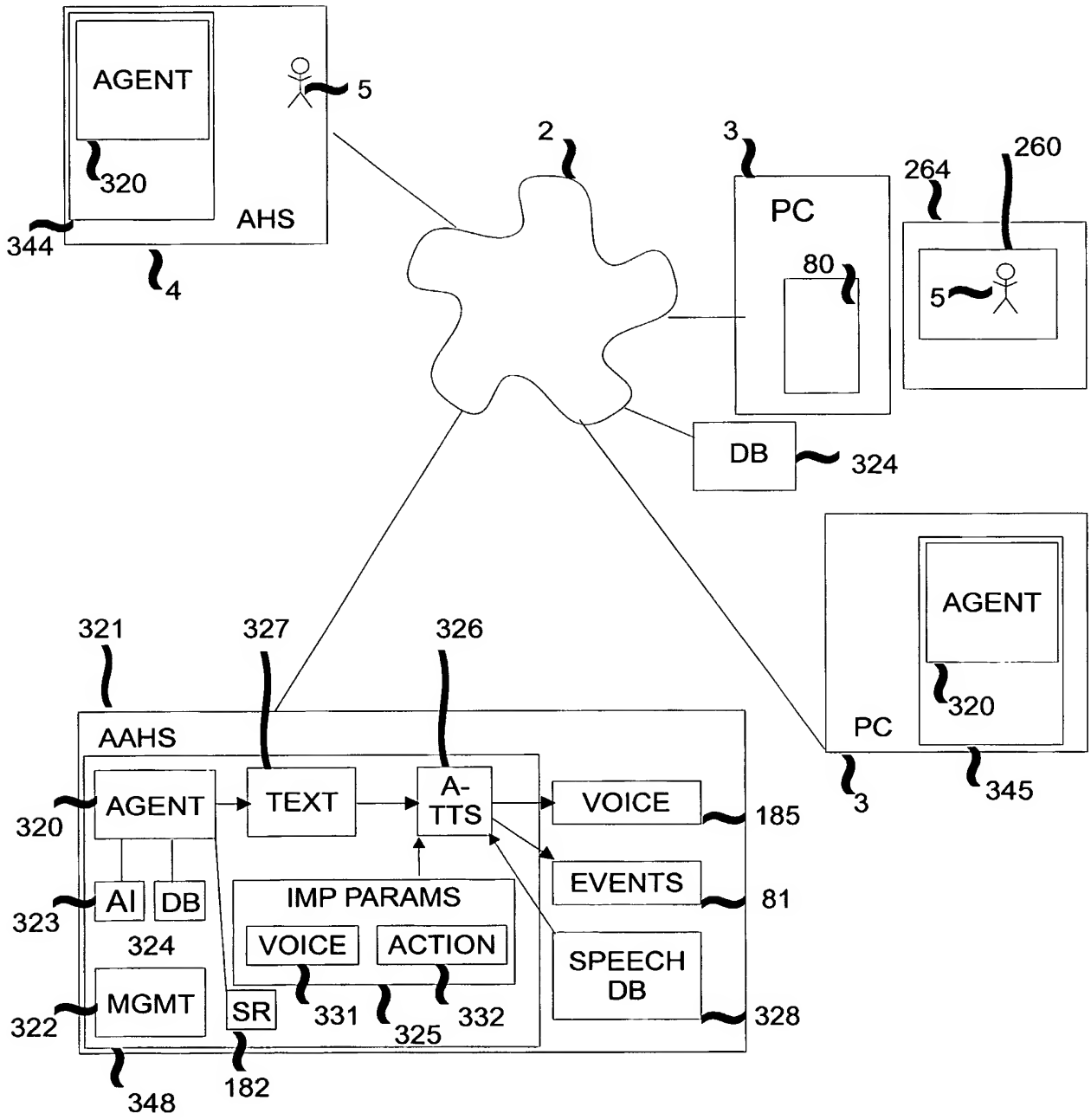

AGENT
320
AHS
344
4
5
2
3
PC
80
264
260
5
DB
324
AGENT
320
PC
345
3
321
327
326
AAHS
AGENT
320
TEXT
A-
TTS
VOICE
185
AI
DB
323
324
IMP PARAMS
VOICE
ACTION
EVENTS
81
MGMT
322
SR
331
325
332
SPEECH
DB
328
348
182

Fig 47

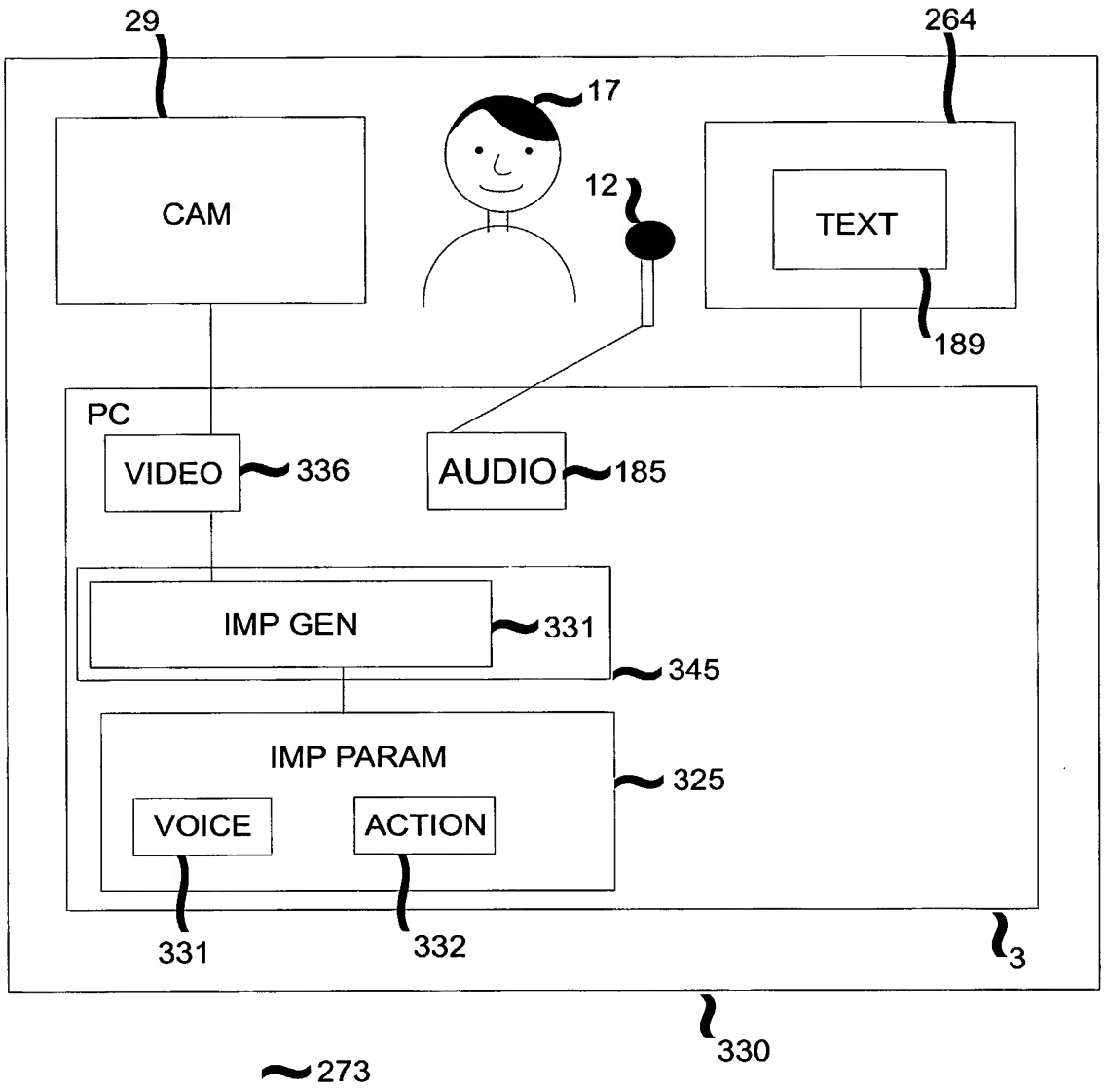

29
264
17
CAM
12
TEXT
189
PC
VIDEO
336
AUDIO
185
IMP GEN
331
345
IMP PARAM
325
VOICE
ACTION
331
332
3
330
273

Fig 48

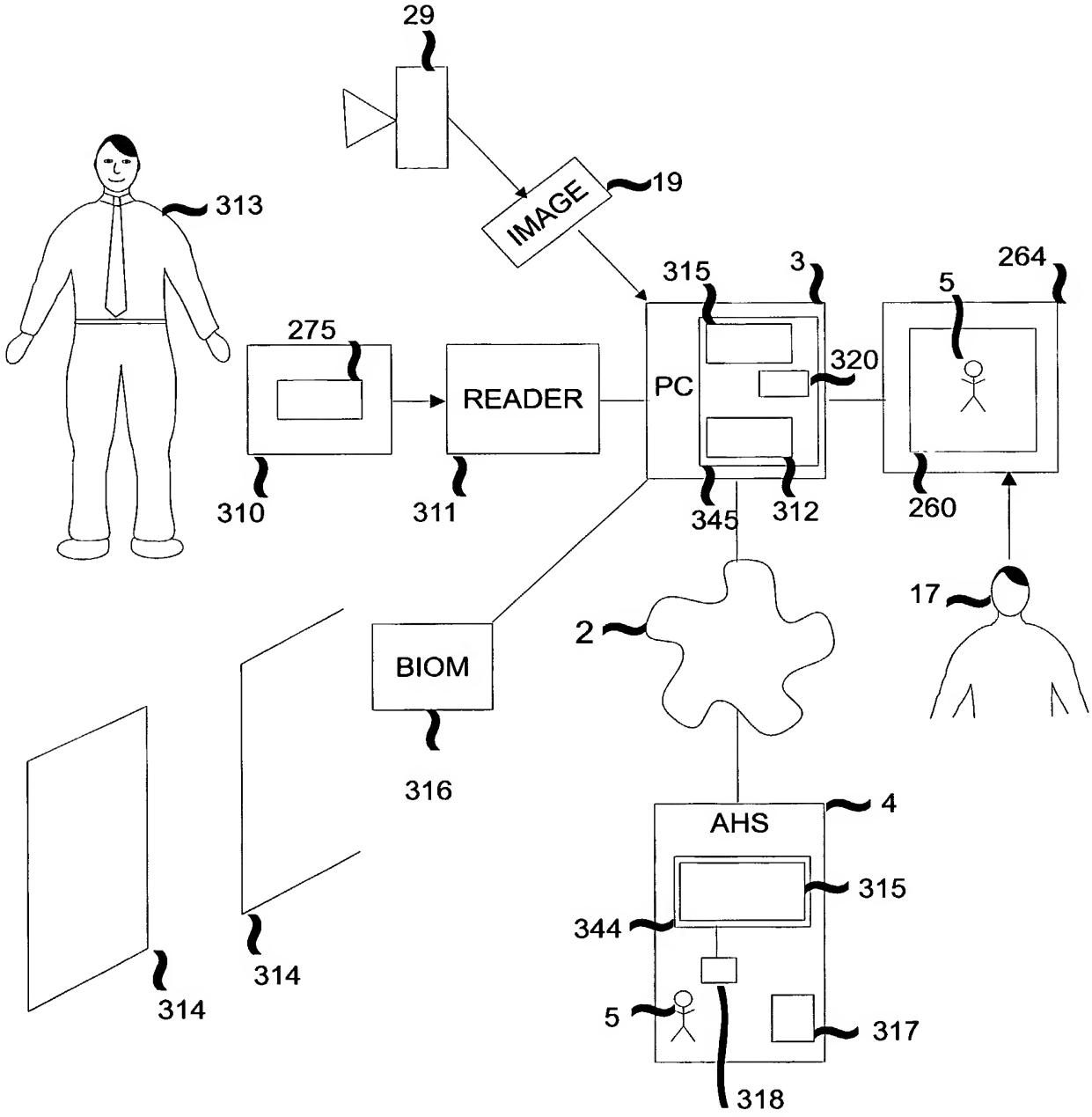

29
IMAGE
19
313
315
3
264
5
275
READER
PC
320
310
311
345
312
260
BIOM
2
17
316
4
AHS
315
344
5
317
318
314
314

Fig 49

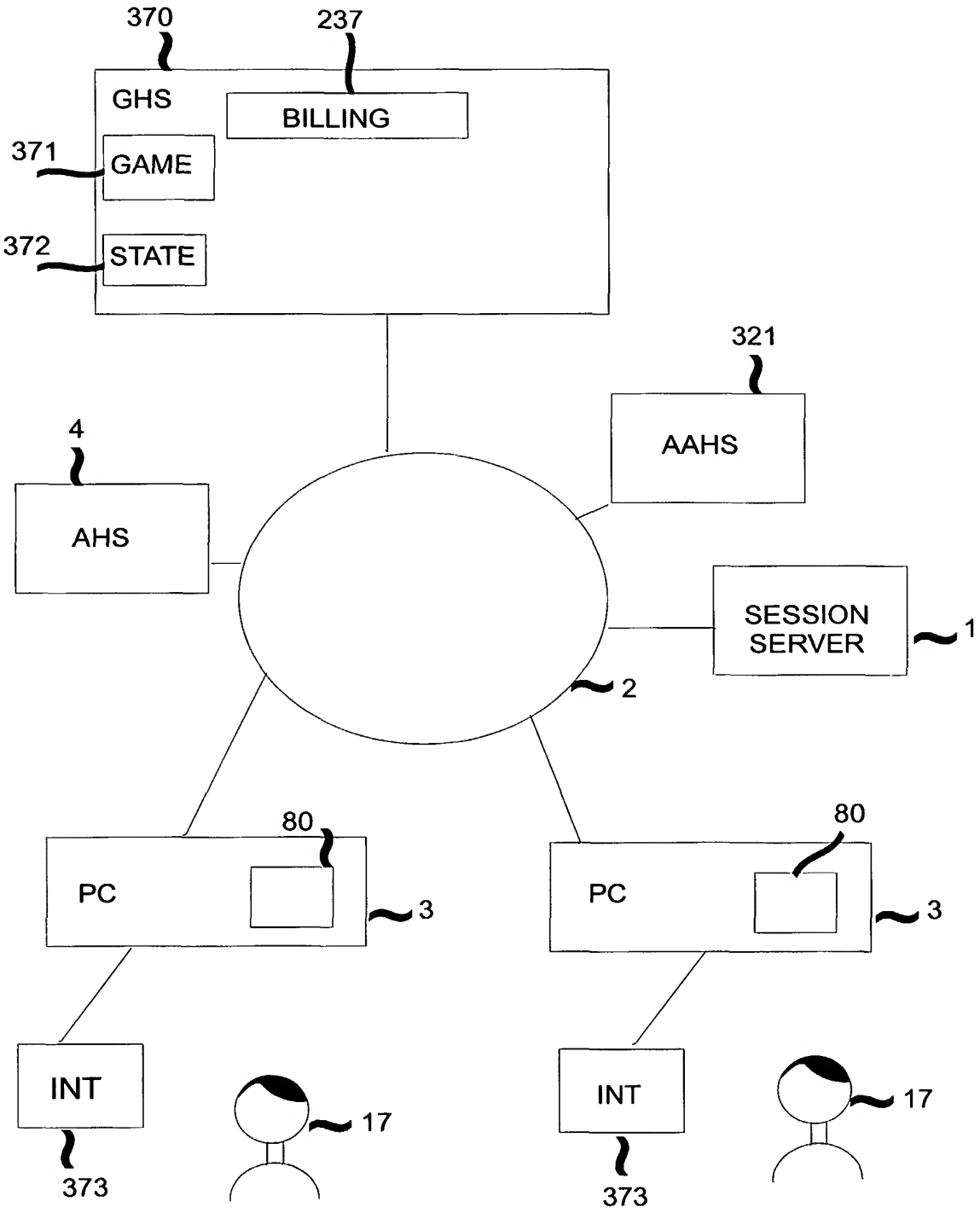

370
237
GHS
BILLING
371
GAME
372
STATE
321
AAHS
4
AHS
SESSION SERVER
1
2
80
PC
3
80
PC
3
INT
373
17
INT
373
17

Fig 50

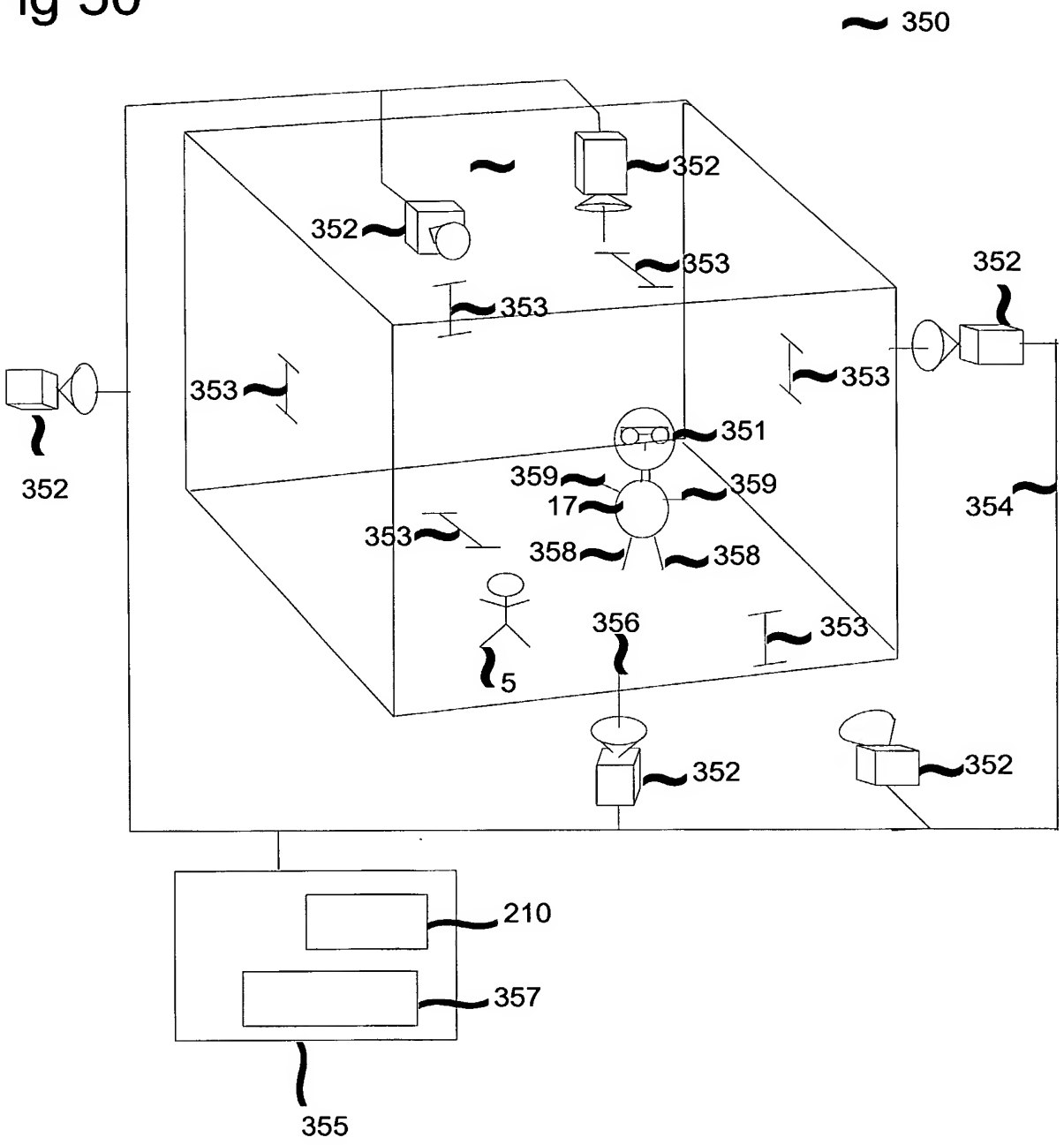

350
352
352
353
353
352
353
353
352
351
359
359
17
354
353
358
358
356
353
5
352
352
210
357
355

Fig 51

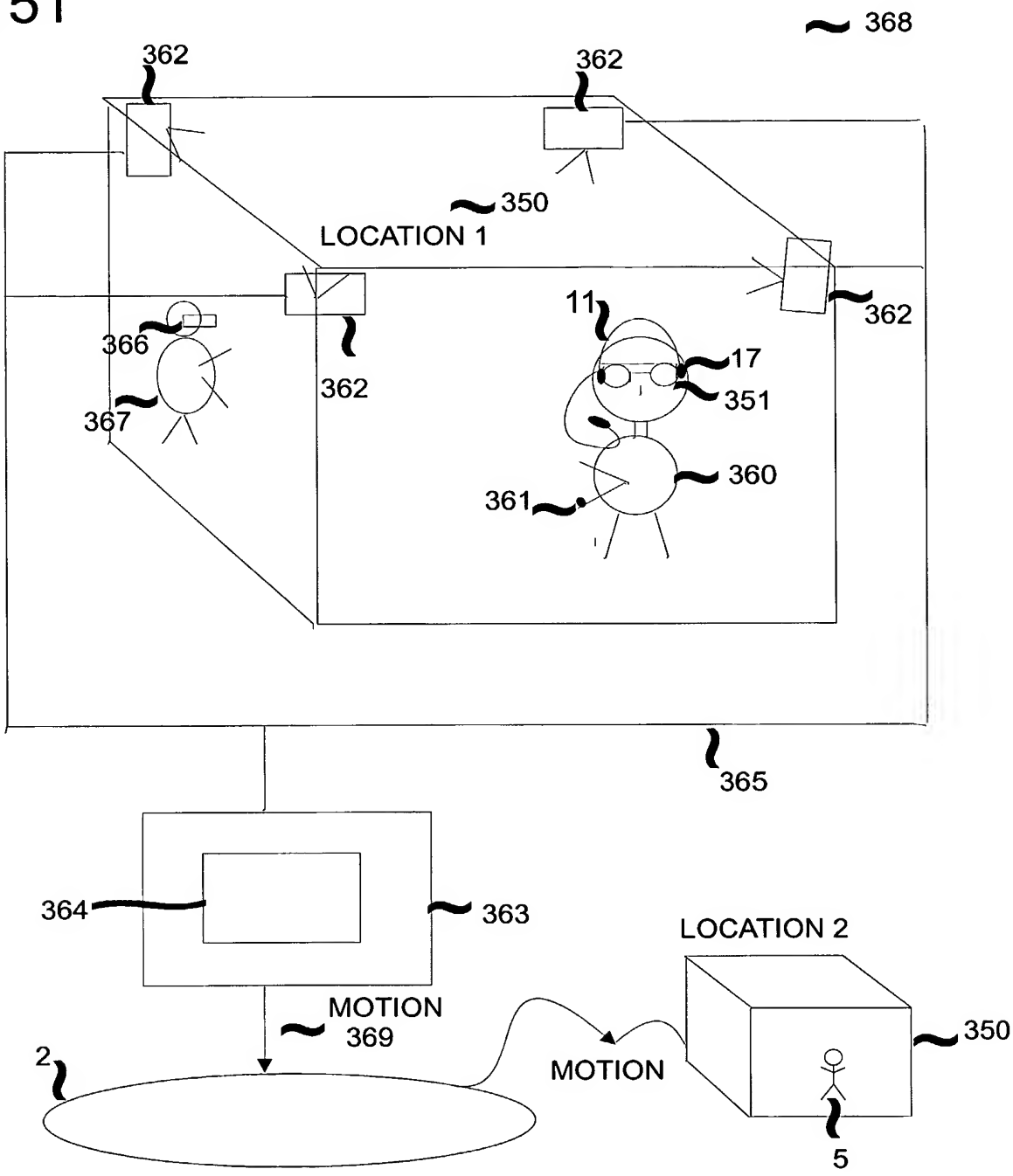

368
362
362
350
LOCATION 1
362
362
366
367
11
17
351
360
361
365
364
363
LOCATION 2
MOTION
369
MOTION
350
2
5

Fig 52

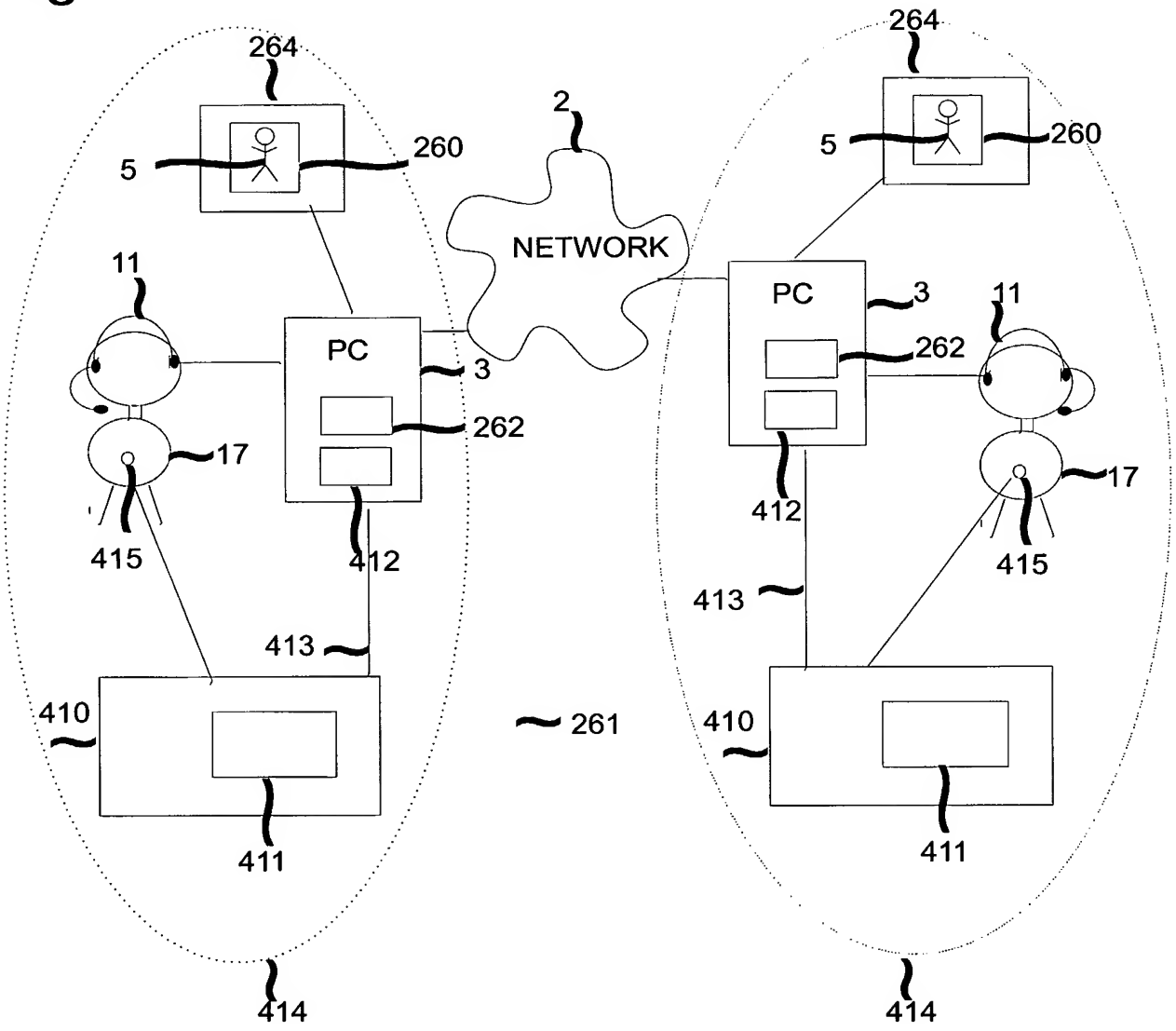

264
5
260
2
NETWORK
11
PC
3
262
17
415
412
413
410
411
261
414
264
5
260
PC
3
11
262
17
412
415
413
410
411
414

Fig 53

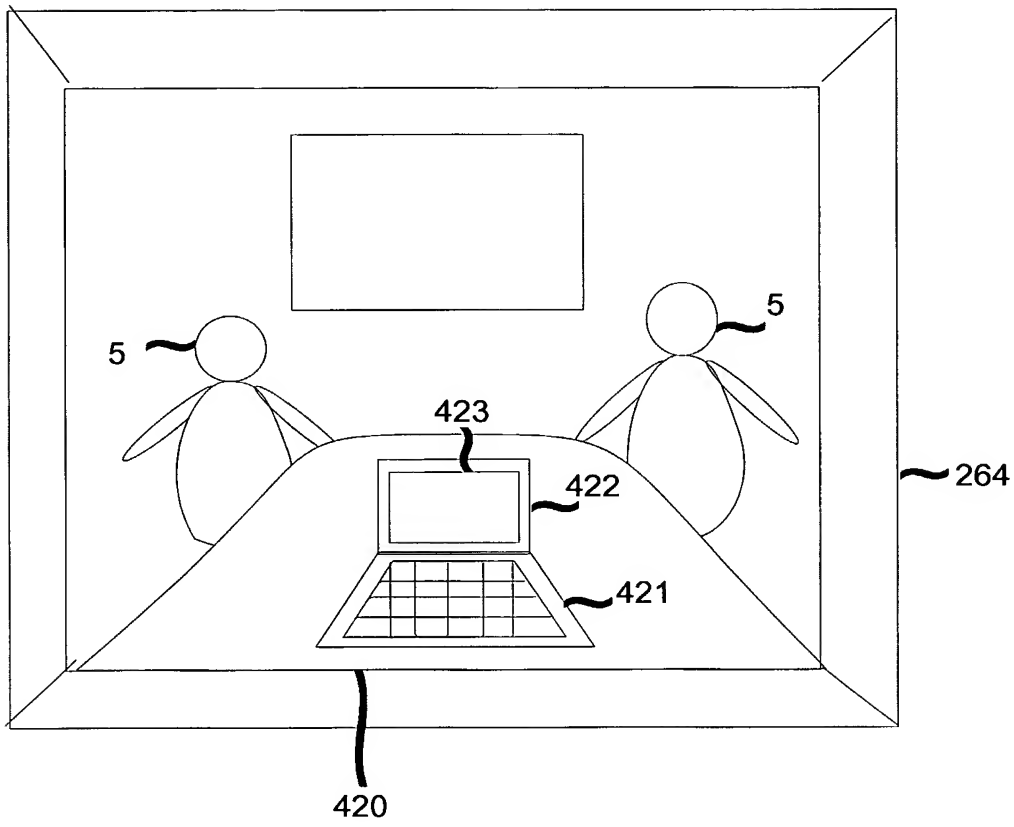


Fig 54

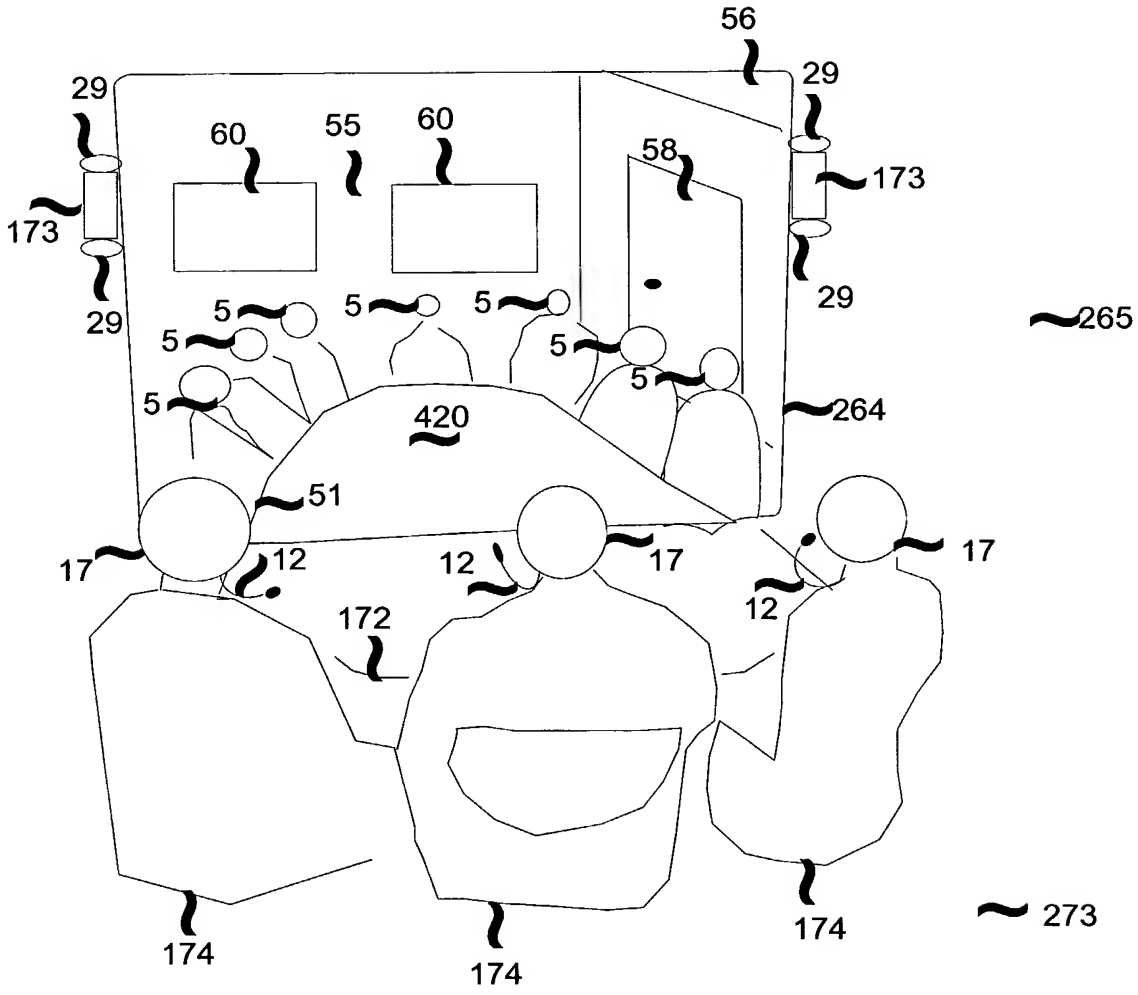

56
29
29
60
55
60
58
173
173
29
29
265
5
5
5
5
5
5
420
264
51
17
12
12
17
12
17
172
174
174
174
273

Fig 55

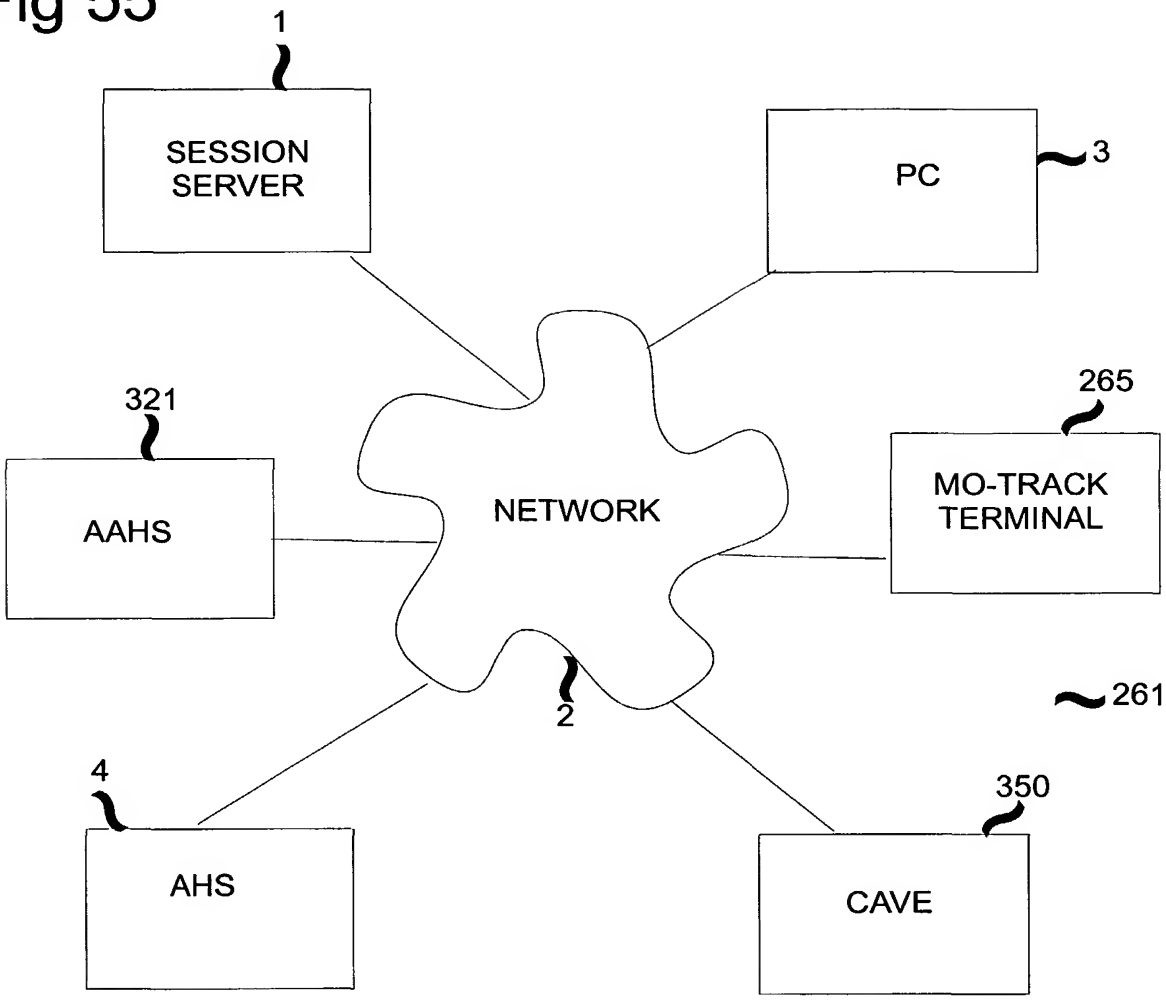

1
SESSION SERVER
PC
3
321
AAHS
NETWORK
265
MO-TRACK TERMINAL
2
261
4
AHS
350
CAVE

Fig 56

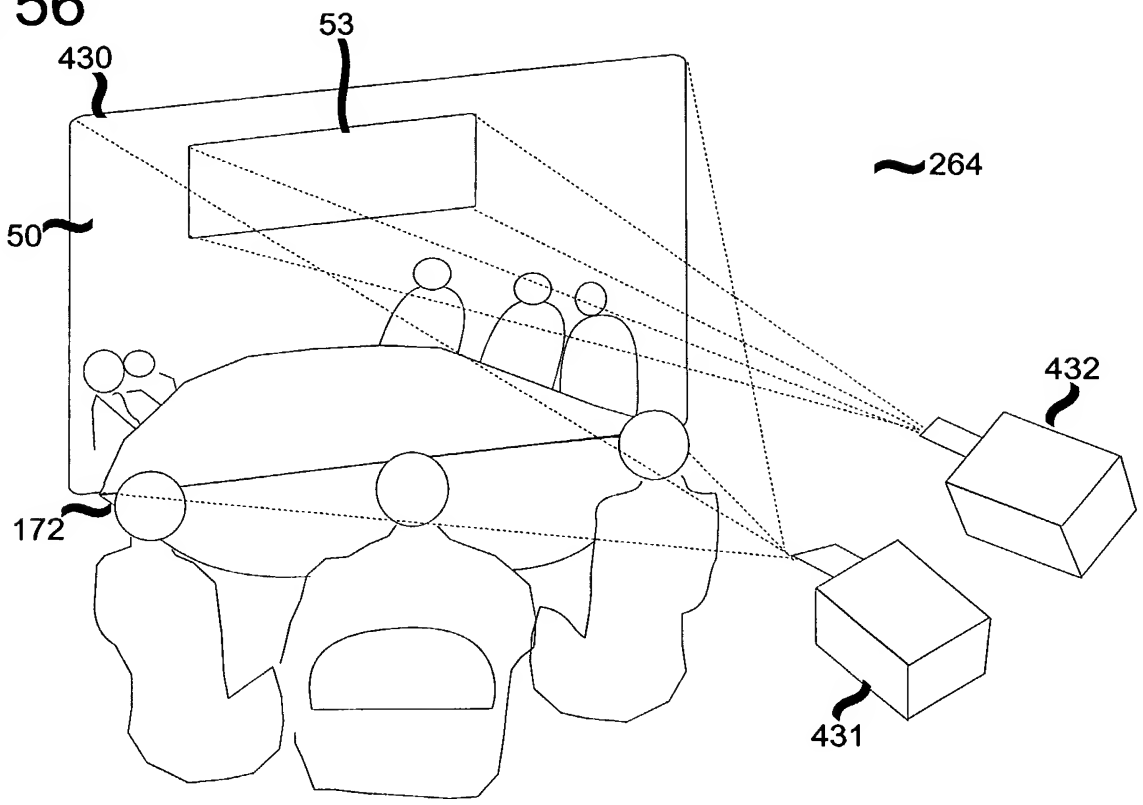


Fig 57

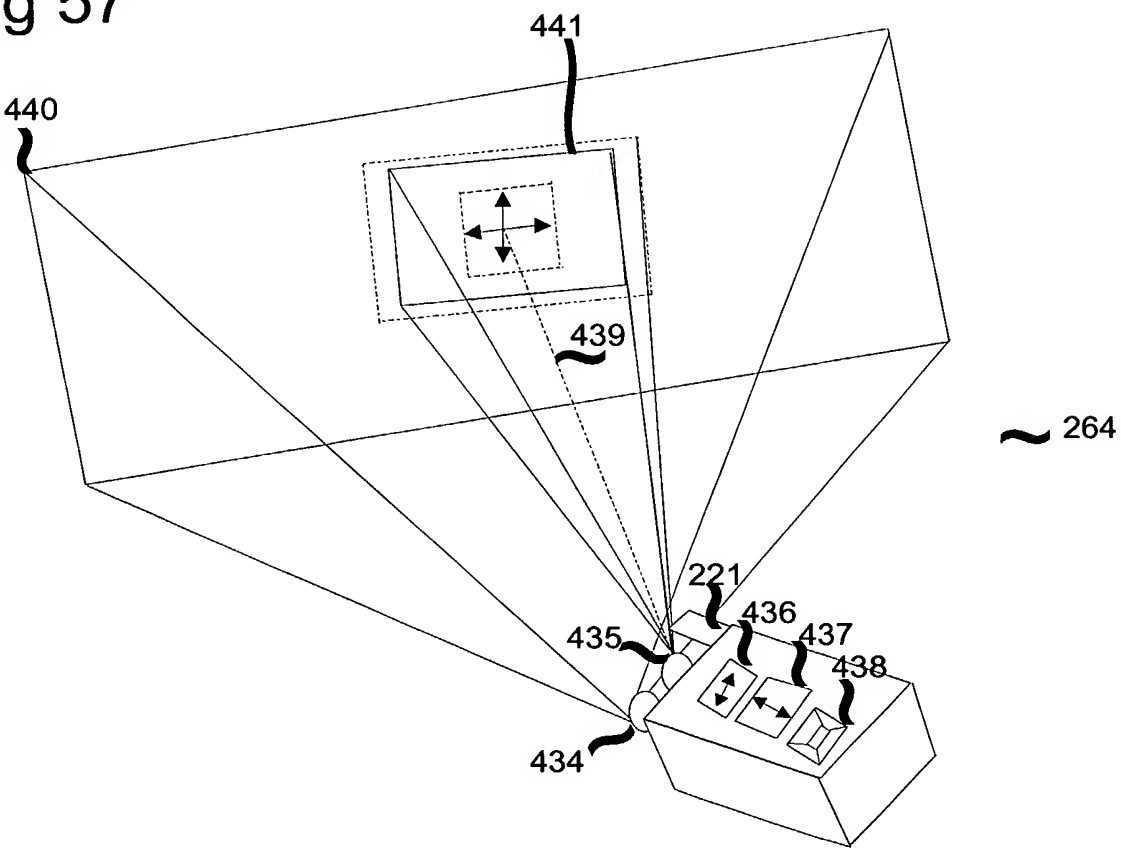


Fig 58

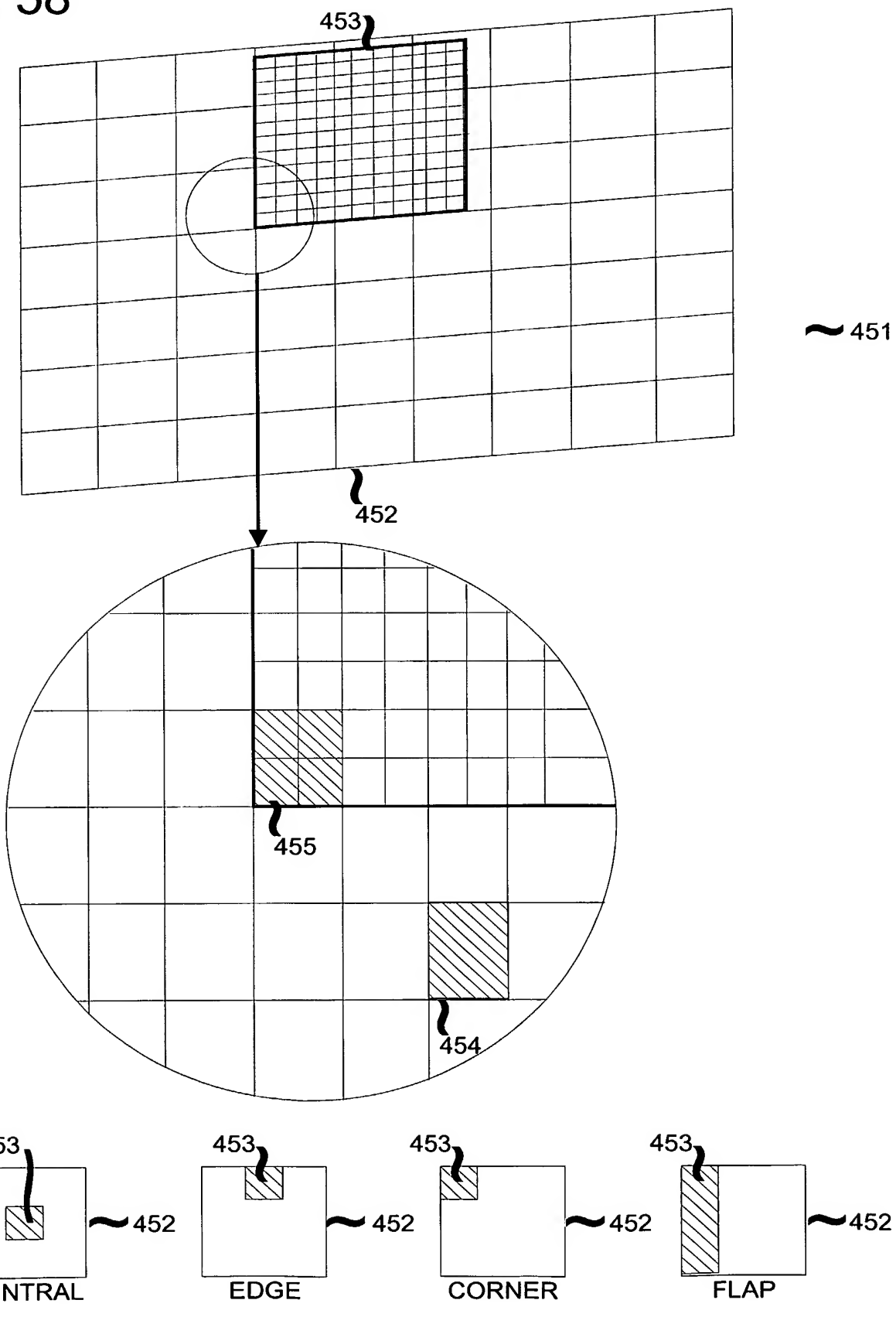

453
451
452
455
454
453
452
CENTRAL
453
452
EDGE
453
452
CORNER
453
452
FLAP

Fig 59

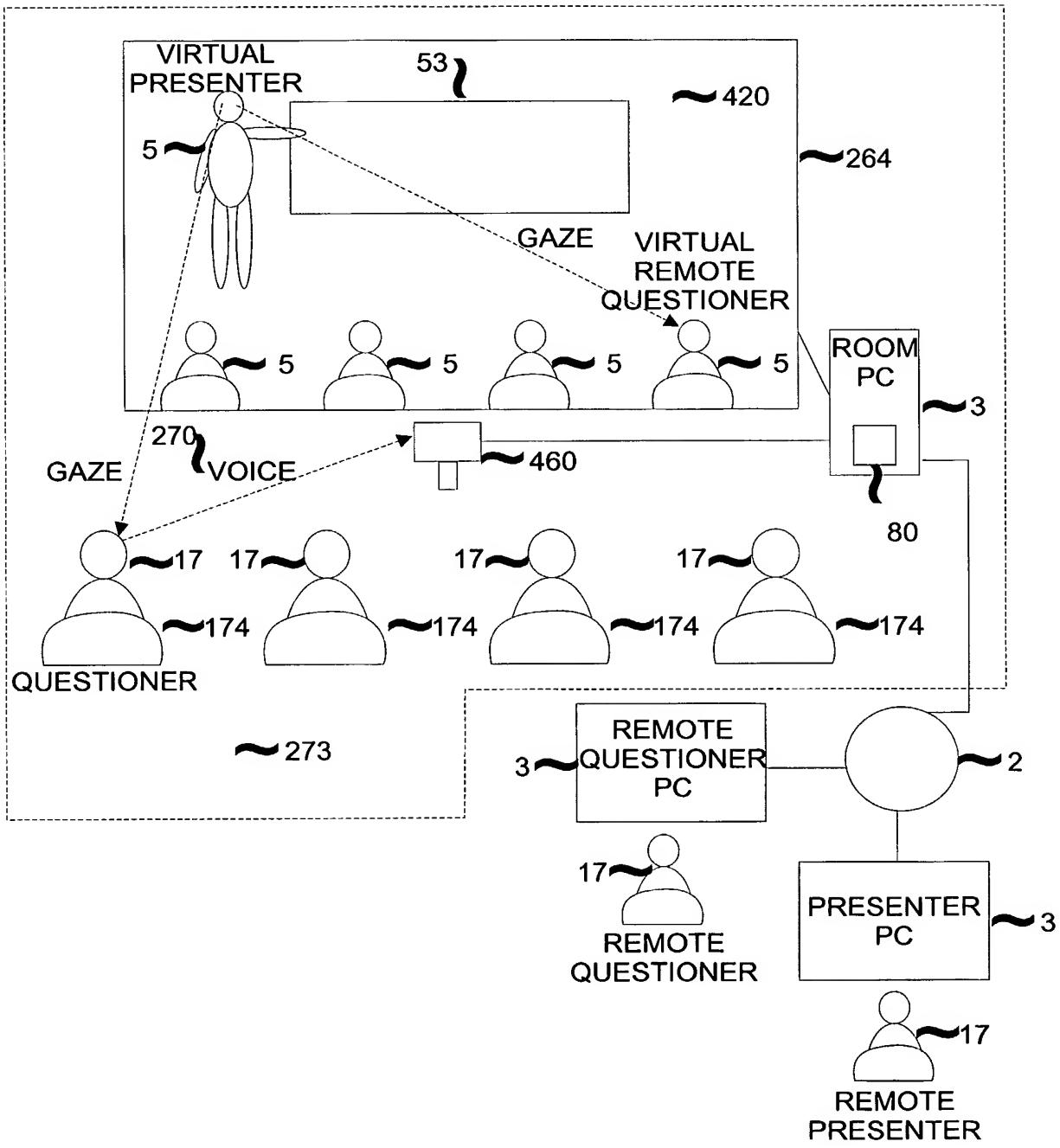
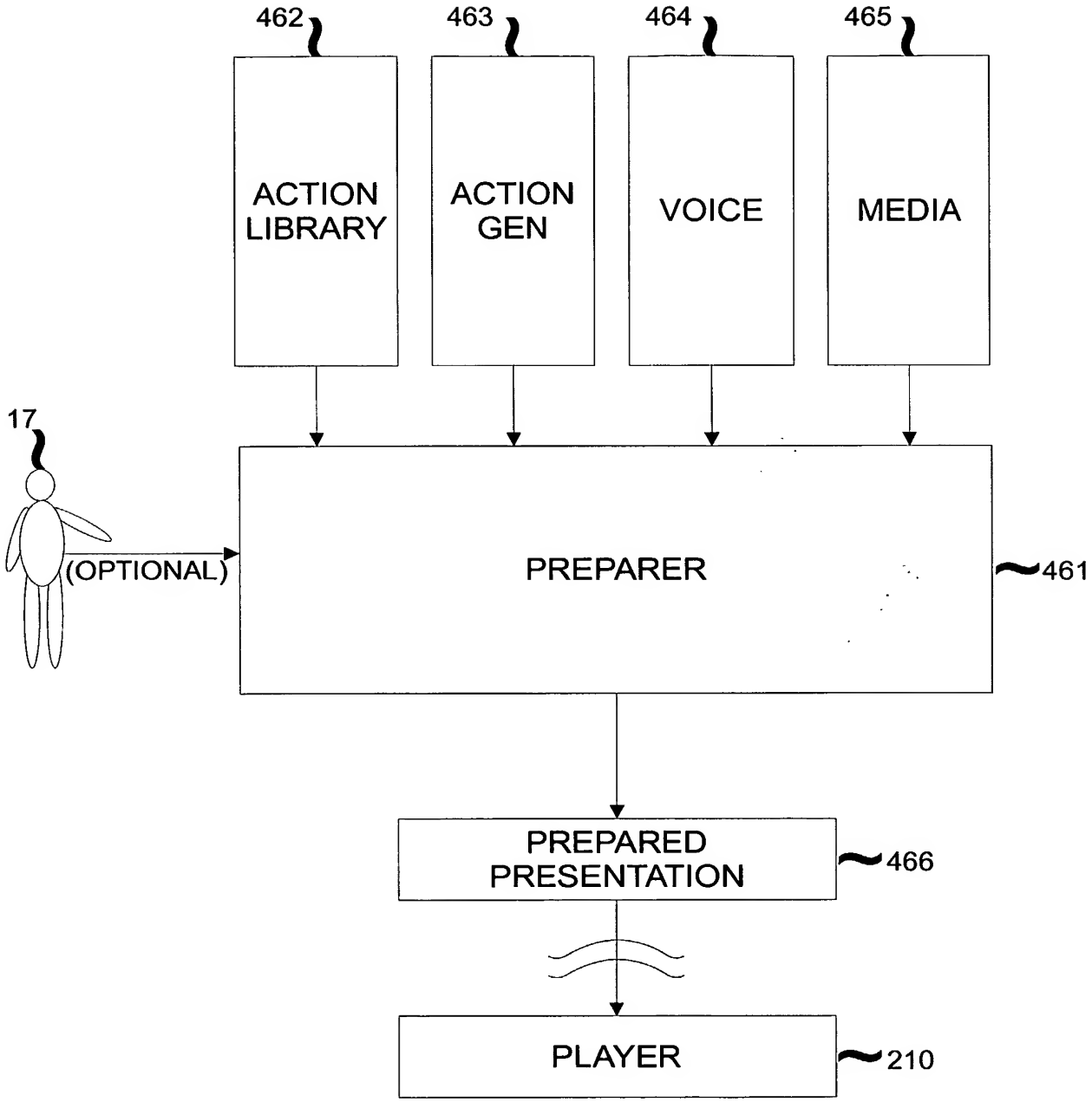

VIRTUAL PRESENTER
53
420
5
264
GAZE
VIRTUAL REMOTE QUESTIONER
5
5
5
5
ROOM PC
3
270
GAZE
VOICE
460
80
17
17
17
17
174
174
174
174
QUESTIONER
273
REMOTE QUESTIONER PC
3
2
17
REMOTE QUESTIONER
PRESENTER PC
3
17
REMOTE PRESENTER

Fig 60


462
463
464
465
ACTION LIBRARY
ACTION GEN
VOICE
MEDIA
17
(OPTIONAL)
PREPARER
461
PREPARED PRESENTATION
466
PLAYER
210